Sequential change point detection in high dimensional time series[*]

Josua Gösmann[1], Christina Stoehr[1], Johannes Heiny[1] and Holger Dette[1]

[1]*Department of Mathematics, Ruhr University Bochum,*
e-mail: [*]josua.goesmann@rub.de; [**]christina.stoehr@rub.de
e-mail: [†]johannes.heiny@rub.de; [‡]holger.dette@rub.de

Abstract: Change point detection in high dimensional data has found considerable interest in recent years. Most of the literature either designs methodology for a retrospective analysis, where the whole sample is already available when the statistical inference begins, or considers online detection schemes controlling the average time until a false alarm. This paper takes a different point of view and develops monitoring schemes for the online scenario, where high dimensional data arrives successively and the goal is to detect changes as fast as possible controlling at the same time the probability of a type I error of a false alarm. We develop a sequential procedure capable of detecting changes in the mean vector of a successively observed high dimensional time series with spatial and temporal dependence. The statistical properties of the method are analyzed in the case where both, the sample size and dimension tend to infinity. In this scenario, it is shown that the new monitoring scheme has asymptotic level alpha under the null hypothesis of no change and is consistent under the alternative of a change in at least one component of the high dimensional mean vector. The approach is based on a new type of monitoring scheme for one-dimensional data which turns out to be often more powerful than the usually used CUSUM and Page-CUSUM methods, and the componentwise statistics are aggregated by the maximum statistic. For the analysis of the asymptotic properties of our monitoring scheme we prove that the range of a Brownian motion on a given interval is in the domain of attraction of the Gumbel distribution, which is a result of independent interest in extreme value theory. The finite sample properties of the new methodology are illustrated by means of a simulation study and in the analysis of a data example.

MSC2020 subject classifications: Primary 62M10, 62H15; secondary 62G20, 60G70.

Keywords and phrases: High dimensional time series, change point analysis, sequential monitoring, Gaussian approximation, bootstrap, extreme value theory.

Received October 2021.

1. Introduction

As digital transformation processes have accelerated during the last decades, new technologies like smartphones or car sensors are able to gather large amounts of data. Due to this development companies, states, research institutes etc. face the problem to manage, monitor and examine huge data sets, which regularly exceed the means of traditional tools. Thus the demand for so-called big data technology is steadily growing and thereby the requirement for theoretical foundation has put a lot of attention at the topic of high dimensional statistics in recent years.

Especially, the topic of change point analysis or detection of structural breaks has regained attraction and numerous authors have started to embed commonly used multivariate methods into a high dimensional framework or even develop new methodology from scratch. Among many others, high dimensional change point problems have been considered by Cho and Fryzlewicz (2015), Wang and Samworth (2018), who develop methodology to identify multiple change points by a (wild) binary segmentation algorithm under sparsity assumptions. Jirak (2015b) and Dette and Gösmann (2018) aggregate component-wise CUSUM-statistics by the maximum functional to detect structural breaks in a sequence of means of a high dimensional time series. Lévy-Leduc and Roueff (2009) analyze internet traffic data, by applying a compenent-wise CUSUM-test to dimension-reduced censored data.

Enikeeva and Harchaoui (2019) employ the Euclidean norm of the CUSUM-process to obtain a linear and a scan statistic of χ^2-type, that is minimax-optimal under the regime of independent Gaussian observations. Change point problems in high dimensional covariance matrices are studied by Wang et al. (2017), Avanesov and Buzun (2018) and Dette et al. (2018) using (wild) binary segmentation, multiscale methods and U-statistics, respectively. U-statistics are also used by Wang et al. (2019) and Wang and Shao (2020) to develop testing and estimation methodology for a structural break in the mean.

All listed references on high dimensional change point problems have in common that the proposed methods are designed for a *retrospective or offline analysis*, where the whole sample is already available when the statistical inference is commenced. In contrast to this, sequential change point detection deals with methods that are applicable for monitoring data in a so-called *online scenario*. In such a setup, data arrives steadily and methods are constructed to detect changes as fast as possible, while the problem is reevaluated with each new data point. Starting with the seminal paper of Wald (1945) on the sequential probability ratio test, an enormous amount of literature has been published discussing the problem from different perspectives. Several concepts have been proposed to model the situation of online monitoring and we discuss the different paradigms addressing the sequential change point problem in more detail at the end of this introduction.

To the authors best knowledge, a common feature of most of the literature on sequential change point detection consists in the fact that it does not consider the time and spatially dependent high dimensional scenario, where the dimen-

sion of the data increases with the sample size. The purpose of the present paper is to address this problem in the context of detecting changes in the mean. For this purpose, we follow the paradigm of Chu et al. (1996), which provides a model for online change point detection, such that the probability of a type I error (false alarm) can be controlled (asymptotically). We develop a sequential algorithm in the high dimensional regime aggregating component-wise sequential detection schemes by the maximum statistic. For the individual components we use a novel monitoring procedure, which screens for all possible positions of the change point and takes into account that the change does not necessarily occur in the first observations after the initial sample – see Section 2 for more details. A nice feature of this approach consists in the fact that the limiting distribution of the statistic used to monitor each component (after appropriate standardization) is given by the *range of the Brownian motion*, that is $\mathbb{M} = \max_{0 \le t \le q} W(t) - \min_{0 \le t \le q} W(t)$, where W is a Brownian motion on the interval $[0, q]$ and $0 < q < 1$ is a known constant. The distribution of the random variable $\mathbb{M}$ appears as the weak limit of the range of cumulative sums of i.i.d. random variables with variance 1 [see Feller (1951)], and we will show that it belongs to the domain of attraction of the Gumbel distribution. This result is of independent interest in extreme value theory and allows us to aggregate component-wise statistics by the maximum.

As a consequence we can develop a sequential monitoring scheme in the high dimensional regime using the quantiles of the Gumbel distribution. For this purpose, we combine Gaussian approximations tools for high dimensional statistics [see Chernozhukov et al. (2013); Zhang and Cheng (2018)] and show that the statistic can be approximated by a counterpart computed from Gaussian observations with the same long-run correlation structure as the observed time series. Using Gaussian comparison and anti-concentration inequalities we show that this statistic is sufficiently close to the maximum of ranges of dependent Brownian motions. Finally, we use Poisson approximation via the Chen-Stein method to eliminate the independence condition, such that the new result for the independent case can be applied. As the rates of most convergence results in extreme value theory are known to be rather slow, we also propose a simple bootstrap procedure, which improves the performance of the sequential monitoring scheme for sample sizes. To our best knowledge, our paper provides the first rigorous analysis of a sequential change point detector in the high dimensional regime for the model introduced by Chu et al. (1996). For the sake of transparency we concentrate on changes in the high-dimensional mean vector, but we expect that our approach can be extended to linear regression models with a high-dimensional parameter. Here the sequential paradigm proposed by Chu et al. (1996) has been considered by Horváth et al. (2004) and Dette and Gösmann (2019) for fixed dimension, among others (see also Kaul et al., 2019; Wang et al., 2021, for some results on retrospective change point analysis in the high dimensional regime).

The remaining part of this paper is organized as follows. We conclude this introduction with a discussion and comparison of the different philosophies in sequential change point detection. In Section 2, we introduce the specific testing

problem under consideration and present the new monitoring procedure for structural breaks in the sequence of means from a high dimensional time series. Section 3 is devoted to our main results and to the analysis of the asymptotic properties of the new procedure. In particular, we prove that the maximum of the individual test statistics converges weakly (with increasing dimension and initial sample size) to a Gumbel distribution. These results are used to show that the monitoring scheme has asymptotic level α and is consistent. In Section 4, we investigate the finite sample properties of the new procedures by means of a simulation study and illustrate potential applications in a data example. Finally, all proofs are deferred to Appendix A in the online supplement.

Related literature – two paradigms in sequential change point detection

In the remaining part of this section, we briefly discuss two ways of modelling the problem of sequential change point detection, which have developed rather independently in the last 25 years. The list of references cited below is by no means complete, because both modelling approaches have been used intensively in the literature. Rather it represents a selective choice of the authors with the goal to put the approach proposed in this paper in the appropriate context. Sequential change point detection in the high dimensional scenario, where the dimension increases with the sample size, has barely been studied in the liteature, where some of the relevant exceptions are cited below.

The different models in the literature are used to address different criteria for quantifying the propensity of a detection scheme to cause false alarms. The traditional *Statistical Process Control* (SPC)-approach has its focus on a fast change point detection, where the average time from monitoring start to a false alarms is controlled.

On the other hand, if the costs of a false alarm are heavy, for instance if a large portfolio has to be restructured in portfolio management, it might be more reasonable to control the probability of a false alarm. While from an application point of view the criteria are strongly related, different mathematical models have to be used to analyze the statistical properties. The SPC-approach uses models from classical sequential analysis (often independent observations) and some of the relevant literature is given in part (B) of this section. As an alternative Chu et al. (1996) introduced a model to control asymptotically the probability of a false alarm, which requires an initial stable sample. The related literature for this model is described in part (A) of this section. The purpose of the present paper is to develop a monitoring scheme in this model, where the dimension is allowed to grow with the sample size (at a polynomial order).

The SPC- and the approach proposed in Chu et al. (1996) have been developed rather independently during the last two decades and usually a manuscript either focuses on the one or on the other. One reason for this development is their different focus in the sequential change point problem and probably other ones are mathematical reasons. The very recent work of Yu et al. (2020) provides a brief theoretical comparison of both methods in the univariate setting with independent sub-Gaussian observations and also indicates the mathemat-

ical difficulties of such a comparison. The authors argue that (for their specific model) controlling the type I error over a fixed monitoring period only results in minor extra costs for the expected detection delay and *"...suggest that upper bounding the overall type I error might be better."* However, their model does not include a stable initial set, such that no asymptotic investigations can be carried out.

In general, methods developed in the SPC-paradigm often base upon the premise, that certain statistical parameters, for instance the mean before a change, are exactly known from a long-lasting stable period (Phase I). The procedure for the actual monitoring (Phase II) is then calibrated using these *exact parameters*; see Chen et al. (2020) for a recent example in the context of high dimensional means. While such a proceeding might be reasonable in some applications like manufacturing control, in other fields like econometrics, this is unrealistic as a large stable set might be rarely available. In this context, the approach of Chu et al. (1996) might be advantageous, as the size of the stable set is treated as a parameter (sample size) and quantities derived from this set are handled as estimators. The problem of separate Phase I and Phase II has also been discussed in the SPC-literature; see Hawkins et al. (2003) or Ross (2014) among others, who discuss SPC-type procedures updating the initial estimates.

(A) Controlling the type I error Chu et al. (1996) propose a sequential paradigm, such that the type I error can be controlled (asymptotically). It is based on the premise of an initial stable data set, that has to be available before monitoring commences. Their main idea is to invoke invariance principles, such that the probability of a false alarm can be controlled over a possibly infinite monitoring time – asymptotically as the size of the initial set grows. This modelling approach also allows for the estimation of model parameters, such as the (long-run) variance or coefficients in regression models if covariates are present, and therefore may have advantages for applications, where temporal dependence is present. The literature distinguishes between the *open-end* and *closed-end scenario*. A *closed-end scenario* is associated with a fixed endpoint, where monitoring has to be eventually stopped even if no change was detected before. An *open-end scenario* does not postulate an endpoint meaning that monitoring can (theoretically) continue forever if no change is detected.

Since its introduction, this paradigm has found considerable attention in the literature on change point detection. For example, Horváth et al. (2004), Hušková and Koubková (2005) and Aue et al. (2006) consider changes in the parameters of linear models with statistics based on residuals. For independent identically distributed (i.i.d.) data, Kirch (2008) and Hušková and Kirch (2012) propose several bootstrap procedures for sequential change point detection in the mean and in the parameters of linear regression models. A MOSUM-approach, which employs a moving monitoring window in linear models is introduced by Chen and Tian (2010), while Ciuperca (2013) proposes a generalization of the *sequential* CUSUM statistic to non-linear models. Fremdt (2014) uses the so-called Page-CUSUM, which scans for changes through the already available monitoring data and is more efficient to detect later changes than the classical sequential

CUSUM scheme. Hoga (2017) proposes an ℓ_1-norm to detect structural breaks in the mean and variance of a multivariate time series and Dette and Gösmann (2019) develop an amplified scanning method combined with self-normalization. Otto and Breitung (2019) define a Backward CUSUM statistic based on recursive residuals in a linear model. Unifying frameworks are provided in Kirch and Kamgaing (2015) and Kirch and Weber (2018) and a theory based on U-statistic is established in Kirch and Stoehr (2019). We also refer to the recent review of sequential procedures in Section 1 of Anatolyev and Kosenok (2018).

(B) Statistical Process Control The traditional approach, partially known as *Statistical Process Control* (SPC), can be traced back to the seminal papers of Page (1954, 1955). So-called *Control Charts* have been the first methods developed in this area and their typical field of application was quality control for manufacturing processes. In this spirit, these methods are made to guarantee a quick change point detection, for which regular false alarms are necessarily accepted. Thus, procedures developed in the SPC-context are commonly analyzed by two quantities. The average run length (ARL) measures the expected time until a false alarm is raised under the null hypothesis of no change. Its counterpart is the expected detection delay (EDD) necessary to raise an alarm under the alternative. It can be regarded as an analogue to classical hypothesis testing theory, that only one of these key quantities can be optimized while the other is only kept bounded. Since Page (1954) numerous authors have followed this approach, for instance, Hinkley (1971), Moustakides (1986) or Nikiforov (1987), and we delegate the reader to review papers by Woodall and Montgomery (1999) or Lai (2001) and the more recent monograph of Tartakovsky et al. (2014), where the state of the art is discussed. In a discussion paper Mei (2008) demonstrates that the average run length (ARL) might not be an appropriate criterion as a measure of false detection if the observations are dependent. The author and the discussants propose several alternatives.

Although the (basic) SPC-approach is relatively old, many authors have been constantly working on this problem from various perspectives. In particular, there also exist several papers developing online monitoring procedures for new situations arising in the information age. For instance Tartakovsky et al. (2006), Mei (2010), Zou et al. (2015) investigate multi-sensor change point problems and Chu and Chen (2018) propose an algorithm based on nearest neighbour information for non-euclidean data. Most of this work investigates multivariate data, where the dimension is assumed to be fixed, and imposes rather strong assumptions to investigate the ARL and EDD (such that of independent identically distributed observations).

On the other hand, there also exist a few references considering the high dimensional case, where the dimension is large compared to the sample size. Xie and Siegmund (2013) develop a mixture procedure to monitor parallel streams of data for a change point without assuming a spatial structure. Soh and Chandrasekaran (2017) combine a filtered derivative approach with convex optimization to develop a scalable and statistically efficient monitoring scheme for high dimensional sparse signals, while Chen et al. (2020) monitor for mean changes

across coordinates by a sophisticated aggregation technique. These three contributions assume independent normal distributed observations without spatial dependence.

While these assumptions are helpful to obtain some basic understanding of the properties of a procedure, they might be too restrictive in applications, in particular in economics. Our approach differs substantially from these high dimensional frameworks. We use the modelling approach by Chu et al. (1996) to (asymptotically) control the probability of a type I error in the high dimensional regime. Moreover, we neither assume Gaussian nor independent observations to analyze the theoretical properties of the proposed monitoring procedure.

2. Sequential monitoring of high dimensional time series

Let $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ denote a time series of random vectors in $\mathbb{R}^d$ with mean vectors

$$\boldsymbol{\mu}_t := (\mu_{t,1}, \ldots, \mu_{t,d})^\top = \mathbb{E}[\boldsymbol{X}_t] := \mathbb{E}[(X_{t,1}, \ldots, X_{t,d})^\top] \ .$$

We take the sequential point of view and are interested in monitoring for changes in the vectors $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots$. Following Chu et al. (1996) we assume that a historic or initial data set, say $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$, is available, which is known to be mean stable. Starting with observation $\boldsymbol{X}_{m+1}$ we will sequentially test for a change in the mean vector in the monitoring period. The corresponding testing problem is therefore given by the hypotheses

$$H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_m = \boldsymbol{\mu}_{m+1} = \boldsymbol{\mu}_{m+2} = \cdots$$
$$\text{vs. } H_1 : \exists k^* \in \mathbb{N}, \text{ s.t. } \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_m = \cdots = \boldsymbol{\mu}_{m+k^*-1} \neq \boldsymbol{\mu}_{m+k^*} = \cdots . \quad (2.1)$$

In the present paper, we consider a closed-end scenario where the procedure stops after $m + Tm$ observations even if no change has been detected [see Aue et al. (2012), Wied and Galeano (2013) among many others]. The factor T determines the length of the monitoring period compared to the size of the initial training set m and so the hypotheses in (2.1) read as follows

$$H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_m = \boldsymbol{\mu}_{m+1} = \boldsymbol{\mu}_{m+2} = \cdots = \boldsymbol{\mu}_{m+Tm} \quad (2.2)$$
$$\text{vs. } H_1 : \exists k^* \in \{1, \ldots, Tm\}, \text{ s.t. } \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_{m+k^*-1} \neq \boldsymbol{\mu}_{m+k^*} = \cdots = \boldsymbol{\mu}_{m+Tm}.$$

In the following, we will develop a sequential detection scheme which is capable to distinguish between the hypotheses given in (2.2) in a high dimensional setting where the dimension d of the mean vector is increasing with the initial sample size m. To be precise, we denote by

$$\widehat{\mu}_i^j(h) = \frac{1}{j-i+1} \sum_{t=i}^{j} X_{t,h}$$

the estimator of the mean in component $h \in \{1, \ldots, d\}$ from the subsample $X_{i,h}, \ldots, X_{j,h}$. Following Gösmann et al. (2020), we consider the statistic

$$\hat{E}_{m,h}(k) = \max_{j=0}^{k-1} \frac{k-j}{\sqrt{m}\hat{\sigma}_h} \left| \widehat{\mu}_{m+j+1}^{m+k}(h) - \widehat{\mu}_1^{m+j}(h) \right| \quad (2.3)$$

at time point $m + k$ in a single component h, where $\hat{\sigma}_h^2$ denotes an appropriate estimator of the unknown long-run variance

$$\sigma_h^2 = \sum_{t \in \mathbb{Z}} \mathrm{Cov}(X_{0,h}, X_{t,h})$$

in the hth component (explicit conditions for the existence of the long-run variance are given in Section 3). Note that $\hat{E}_{m,h}(k)$ is a weighted CUSUM statistic to detect a change point in the sequence of means corresponding to the data $X_{m+1,h}, \ldots, X_{m+k,h}$. A structural break in the sequence of means $\mu_{m+1,h}, \mu_{m+2,h}, \ldots$ is detected as soon as the sequence

$$w(1/m)\hat{E}_{m,h}(1), w(2/m)\hat{E}_{m,h}(2), \ldots$$

exceeds a given threshold, that is $w(k/m)\hat{E}_{m,h}(k) > c_\alpha^{(h)}$, where w is a suitable weight function and the critical value $c_\alpha^{(h)}$ is chosen based on the desired test level α. Following Aue and Horváth (2004), Wied and Galeano (2013) and Fremdt (2014), we will work with the commonly used weight function $w(t) = 1/(1+t)$, throughout this paper.

Remark 2.1. (1) Note that most of the literature investigates sequential detectors based on the differences

$$\left| \hat{\mu}_{m+1}^{m+k}(h) - \hat{\mu}_1^m(h) \right| \tag{2.4}$$

and the corresponding detection schemes are usually called (ordinary) CUSUM tests [see Chu et al. (1996), Horváth et al. (2004), Aue et al. (2006)]. Another part of the literature focuses on detectors based on the differences

$$\left| \hat{\mu}_{m+j+1}^{m+k}(h) - \hat{\mu}_1^m(h) \right| \text{ for } j = 0, \ldots, k-1 \tag{2.5}$$

and the corresponding detection schemes are usually called Page-CUSUM tests [see Fremdt (2014, 2015), Kirch and Weber (2018)]. The use of the differences $\left| \hat{\mu}_{m+j+1}^{m+k}(h) - \hat{\mu}_1^{m+j}(h) \right|$ is motivated by the likelihood principle [see Dette and Gösmann (2019)]. Compared to the differences in (2.4) it avoids the problem that the estimator $\hat{\mu}_{m+1}^{m+k}(h)$ may be corrupted by observations before the change point, which could lead to a loss of power. Compared to the differences in (2.5) the use of $\hat{\mu}_1^{m+j}(h)$ instead of $\hat{\mu}_1^m(h)$ may avoid a loss in power in cases of a small initial sample and a rather late change point. The advantages of detection schemes based on the differences $\left| \hat{\mu}_{m+j+1}^{m+k}(h) - \hat{\mu}_1^{m+j}(h) \right|$ against ordinary sequential CUSUM and the Page-CUSUM procedures have been recently demonstrated by Gösmann et al. (2020).

(2) Several authors consider the more general class of weight functions $w_\gamma(t) = (t+1)^{-1}\left(\frac{t}{t+1}\right)^{-\gamma}$ for $\gamma \in [0, 1/2)$ [see for instance Horváth et al. (2004), Aue et al. (2006) or Kirch and Weber (2018)]. The weight function $w(t) =$

$1/(1 + t)$ is obtained for $\gamma = 0$ and was proven to be preferable to $\gamma > 0$ in many situations except for changes that occur almost immediately [see Kirch and Weber (2018)]. It is most likely, that the theoretical results of this paper remain correct in the case $\gamma > 0$.

(3) Note that the individual statistics $w(k/m)\hat{E}_{m,h}(k)$ correspond to the special case of Gösmann et al. (2020) with $\gamma = 0$ in dimension one. As further the mean functional μ is linear, straightforward calculations show that its running maximum until time $m + t$ can be equivalently written as

$$\max_{k=1}^{t} w(k/m)\hat{E}_{m,h}(k) = \frac{\sqrt{m}}{\hat{\sigma}_h}\Big(\max_{k=0}^{t}\hat{\mu}_1^{m+k}(h) - \min_{k=0}^{t}\hat{\mu}_1^{m+k}(h)\Big) \,,$$

which is in some sense simpler and less computationally demanding. Nevertheless, we stick with the original representation from (2.3), which – in the authors' view – is more intuitive due to the observable data separation. Further, it offers some substantial advantages in the proofs regarding the truncation and Gaussian approximation steps as well as for the proof of consistency under the alternative.

In order to control the probability of erroneously deciding for a structural break in the component h during the monitoring period, one has to determine the probability

$$\mathbb{P}\Big(\max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h}(k) > c_\alpha^{(h)}\Big) \,.$$

For fixed $h \in \{1, \ldots, d\}$ we can employ a result of Gösmann et al. (2020) who showed that (under appropriate assumptions), as $m \to \infty$,

$$\max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h}(k) \overset{\mathcal{D}}{\Longrightarrow} \mathbb{M} = \max_{0 \leq t \leq q(T)} W(t) - \min_{0 \leq t \leq q(T)} W(t) \,, \qquad (2.6)$$

where the symbol $\overset{\mathcal{D}}{\Longrightarrow}$ denotes weak convergence, $q(T) = T/(T + 1)$ and W is a standard one-dimensional Brownian motion. Note that T is the parameter controlling the length of the monitoring period [see the hypotheses in (2.2)]. $\mathbb{M}$ is known in the probability literature as the range of the Brownian motion on the interval $[0, q(T)]$ and its distribution appears as the weak limit of the range of cumulative sums of i.i.d. random variables with variance 1 [see Feller (1951)].

For a detection of a change point in the complete mean vector we propose to aggregate the statistics for the different spatial dimensions $h = 1, \ldots, d$. More precisely, we consider the maximum of the different components, that is $\max_{h=1}^{d} w(k/m)\hat{E}_{m,h}(k)$, and reject the (closed-end) null hypothesis H_0 : $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_{m+Tm}$ of no structural break in the high dimensional means $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{m+Tm}$ if this quantity exceeds a given threshold, that is

$$\widehat{\mathcal{T}}_{m,d} := \max_{k=1}^{Tm} \max_{h=1}^{d} w(k/m)\hat{E}_{m,h}(k) > c_{d,\alpha} \,. \qquad (2.7)$$

Here the critical value $c_{d,\alpha}$ is chosen appropriately such that (asymptotically) the probability of erroneously deciding for a change point is controlled. In the

following section, we investigate the weak convergence of the statistic $\widehat{\mathcal{T}}_{m,d}$. These results will be used to define critical values $c_{d,\alpha}$ in (2.7) (one by asymptotic theory and one by bootstrap), such that the monitoring procedure is consistent and at the same time controls the probability of the type I error, that is

$$\limsup_{m,d\to\infty} \mathbb{P}_{H_0}\left(\widehat{\mathcal{T}}_{m,d} > c_{d,\alpha}\right) \leq \alpha \quad \text{and} \quad \lim_{m,d\to\infty} \mathbb{P}_{H_1}\left(\widehat{\mathcal{T}}_{m,d} > c_{d,\alpha}\right) = 1 . \quad (2.8)$$

Remark 2.2. In this paper, we consider mean based detectors corresponding to ordinary least squares estimation. We expect that similar results can be obtained using robust estimates such as M-estimates considered by Chochola et al. (2013). For example, the median has been discussed as a special case in Dette and Gösmann (2019) and Gösmann et al. (2020).

3. Main results

In this section, we derive the asymptotic properties of the proposed detector defined in (2.7) in the high dimensional setting where sample size and dimension tend to infinity and we allow for temporal as well as spatial dependencies in the data. In particular, we establish as a consequence of Theorem 3.6 below – in case of constant mean vectors – the weak convergence

$$a_d\left(\widehat{\mathcal{T}}_{m,d} - b_d\right) \overset{\mathcal{D}}{\Longrightarrow} G \quad \text{as } m,d\to\infty ,$$

where a_d, b_d are suitable sequences and G is a standard Gumbel random variable with c.d.f. $F_G(x) = \exp(-\exp(-x))$, $x \in \mathbb{R}$. As inevitable in high dimensional time series analysis, we require assumptions on the relation between the (initial) sample size and the dimension as well as assumptions on the dependence structure to control the dependence between components at different time points uniformly.

Throughout this paper, we assume that the observations are drawn from the array $\{X_{t,h}\}_{t\in\mathbb{Z},h\in\mathbb{N}}$, for which we suppose the location model

$$X_{t,h} = \mu_{t,h} + e_{t,h} , \quad t \in \mathbb{Z}, \ h \in \mathbb{N} , \quad (3.1)$$

where $\mu_{t,h} = \mathbb{E}[X_{t,h}]$ is the expectation of the h component and the centered array $\{e_{t,h}\}_{t\in\mathbb{Z},h\in\mathbb{N}}$ is given as a physical system [see e.g. Wu (2005)], that is

$$e_{t,h} = g_h(\varepsilon_t, \varepsilon_{t-1}, \dots) , \quad t \in \mathbb{Z}, \ h \in \mathbb{N} . \quad (3.2)$$

The underlying sequence of innovations $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ consists of i.i.d. random variables with values in some arbitrary measure space $\mathbb{S}$ and the functions $g_h : \mathbb{S}^{\mathbb{N}} :\to \mathbb{R}$ are assumed to be measurable. Note that, by the definition above, the random variables $\{e_{t,h}\}_{t\in\mathbb{Z},h\in\mathbb{N}}$ are (strictly) stationary with respect to the time index t, such that for any fixed dimension d the multivariate time series $\left\{\boldsymbol{e}_t = \left(e_{t,1}, e_{t,2}, e_{t,3}, \dots, e_{t,d}\right)^\top\right\}_{t\in\mathbb{Z}}$ is stationary. The data generating model defined by formula (3.2) has received a lot of attention in recent years [see for

example Wu and Zhou (2011), Liu et al. (2013), El Machkouri et al. (2013), Berkes et al. (2014) among many others]. It covers the major part of prevalent time series models like autoregressive or moving average processes. Furthermore, it also allows for a natural measurement of temporal dependence which is constructed as follows. Let ε_0' be an independent copy of ε_0 and define

$$X_{t,h}' = \mu_{t,h} + g_h(\varepsilon_t, \varepsilon_{t-1}, \dots \varepsilon_1, \varepsilon_0', \varepsilon_{-1}, \dots)$$

as a counterpart of $X_{t,h}$ where ε_0 is replaced by ε_0'. If $p \geq 1$ we denote by $\|X\|_p = \mathbb{E}[|X|^p]^{1/p}$ the ordinary L_p-norm of a real-valued random variable X (assuming its existence). If $\|e_{t,h}\|_p < \infty$, the coefficients

$$\vartheta_{t,h,p} := \left\| X_{t,h} - X_{t,h}' \right\|_p$$

measure the influence of innovation ε_0 on $X_{t,h}$ and thereby quantify the (temporal) dependence within the system $\{e_{t,h}\}_{t\in\mathbb{Z}, h\in\mathbb{N}}$ defined by (3.2). If $\|e_{t,h}\|_p < \infty$ for some $p \geq 2$ we define the covariances of cross-components by $\phi_{t,h_1,h_2} := \mathrm{Cov}(X_{0,h_1}, X_{t,h_2}), \phi_{t,h} := \phi_{t,h,h}$ and the long-run covariances and variances by

$$\gamma_{h_1,h_2} := \sum_{t\in\mathbb{Z}} \phi_{t,h_1,h_2} \quad \text{and} \quad \sigma_h^2 := \gamma_{h,h}\,, \tag{3.3}$$

respectively. If $\sigma_{h_1}, \sigma_{h_2} > 0$, let additionally

$$\rho_{h_1,h_2} := \frac{\gamma_{h_1,h_2}}{\sigma_{h_1}\sigma_{h_2}} \tag{3.4}$$

denote the long-run correlations. It will be crucial for the asymptotic considerations to control the coefficients $\vartheta_{t,h,p}$ and the correlations ρ_{h_1,h_2} for increasing time t and spatial distance $|h_1 - h_2|$, respectively. This will be formulated in Assumptions 3.3 and 3.4 below. Before we state these precisely, we begin with two assumptions on the relation between sample size and dimension and on the tail behavior of the errors in model (3.1).

Assumption 3.1 (Assumption on the dimension) There exist constants $D > 0$ and $C_D > 0$, such that

(D1) $m^{1/C_D} \leq d \leq C_D\, m^D$.

Assumption 3.2 (Structural assumptions) The random variables $e_{t,h}$ in model (3.1) have bounded exponential moments, that is: there exists a positive sequence B_m, such that

(S1) $\max_{h=1}^d \mathbb{E}\big[\exp\big(|e_{1,h}|/B_m\big)\big] \leq C_e\,,$

where $B_m \leq m^B$ with constants $B < 3/8$ and $C_e > 1$.

Assumption 3.3 (Temporal dependence) There exist constants $p > 2D + 4$, $\beta \in [0,1)$, $C_\vartheta > 0$ such that for all $t \in \mathbb{N}_0$

(TD1) $\sup_{h\in\mathbb{N}} \vartheta_{t,h,p} \leq C_\vartheta \beta^t$.

Further, assume that for a positive constant c_σ the long-run variances defined in (3.3) are uniformly bounded from below, that is

(TD2) $c_\sigma \leq \inf_{h \in \mathbb{N}} \sigma_h$.

Assumption 3.4 (Spatial dependence) There exist a sequence r_m converging to zero, and a constant $\rho_+ \in [0, 1)$, such that the long-run correlations defined in (3.4) fulfill

(SD1) $|\rho_{i,j}| \leq \big(\log |i - j| \big)^{-2} r_{|i-j|}$ whenever $|i - j| \geq 2$,

(SD2) $\sup_{i,j:\, |i-j| \geq 1} |\rho_{i,j}| \leq \rho_+ < 1$.

Let us briefly discuss the assumptions above. Assumptions of the type (D1) are quite common in high dimensional change point problems. For example, Jirak (2015b), Wang and Samworth (2018) and Dette and Gösmann (2018) also assume a polynomial upper bound on the growth of the dimension with the sample size. The lower bound in (D1) is needed for an intermediate truncation step in our proofs where we employ the Fuk–Nagaev inequality in order to obtain the Gaussian approximation. Conditions like Assumption 3.2 and 3.3 are both indispensable ingredients for Gaussian approximation results in high dimensional statistics and will be used in the proofs of our main results. Here Assumption 3.2 controls the tail behavior of the observations [see also Chernozhukov et al. (2013, 2019)], while Assumption 3.3 states a sufficiently weak temporal dependence and a lower bound for the (long-run) variances [see Zhang and Cheng (2018)]. Assumption 3.4 controls the long-run correlations between different components. The fact that the correlations $\rho_{i,j}$ are sufficiently small for a large distance $|i-j|$ is crucial to obtain the desired extreme value convergence. Both parts, (SD1) and (SD2), are in line with those proposed for the retrospective change point method of Jirak (2015b). We point out, that the results of this paper remain correct in the case, where condition (SD1) in Assumption 3.4 is only satisfied after an appropriate permutation of the spatial components. This feature reflects the fact, that the proposed maximum aggregation is invariant with respect to the order of the components.

3.1. Monitoring using asymptotic quantiles

In this section, we develop the asymptotic theory to define a quantile in the monitoring scheme (2.7). Our first result provides the basis for the proof of the main Theorem 3.6 of this section. It is stated here because of its independent interest. It shows that the distribution of the maximum of dependent copies of the random variable $\mathbb{M}$ in (2.6) is in the domain of attraction of the Gumbel distribution if the dependence structure is sufficiently weak.

Theorem 3.5 *Let $(W_1, \ldots, W_d)^\top$ be a d-dimensional Brownian motion with correlation matrix $\Sigma_d = (\rho_{i,j}^{(d)})_{1 \leq i,j \leq d}$ whose entries satisfy $\sup_{1 \leq i < j \leq d} |\rho_{i,j}^{(d)}| \leq \rho_+ \in [0, 1)$. Assume that there exists a sequence L_d, such that $\rho_{i,j}^{(d)} = 0$ if $|i-j| >$*

L_d and

$$L_d = o(d^\Delta) \ \text{for some} \ \Delta < \left(\frac{1 - \rho_+ \sqrt{2 - \rho_+^2}}{1 - \rho_+^2} \right)^2 \ \text{as} \ d \to \infty \,. \qquad (3.5)$$

Further, denote for $q \in (0, 1]$ and $h = 1, \ldots, d$ by $M_h = \max_{0 \le t \le q} W_h(t) - \min_{0 \le t \le q} W_h(t)$ the range of the Brownian motion W_h in the interval $[0, q]$. Then we obtain for $d \to \infty$

$$a_d \left(\max_{h=1}^{d} M_h - b_d \right) \overset{\mathcal{D}}{\Longrightarrow} G \,,$$

where G denotes a standard Gumbel distributed random variable with cdf $F_G(x) = \exp(-\exp(-x))$. The scaling sequences a_d, b_d are given by

$$a_d = \sqrt{\frac{2 \log d}{q}} \quad \text{and} \quad b_d = \sqrt{2q \log d} - \frac{\sqrt{q} \left(\log \log d - \log \frac{16}{\pi} \right)}{2\sqrt{2 \log d}} \,. \qquad (3.6)$$

In the proof of Theorem 3.6 below, we will use a Gaussian approximation which leads to the maximum of the ranges M_h such that Theorem 3.5 can be applied. As indicated by (2.8), the limit distribution of the statistic $\widehat{\mathcal{T}}_{m,d}$ defined in (2.7) has to be derived for the case $m, d \to \infty$ in order to determine an appropriate asymptotic critical value. For this purpose, recall the definition of $\hat{E}_{m,h}(k)$ in (2.3) and define by

$$\begin{aligned}
\mathcal{T}_{m,d} &:= \max_{h=1}^{d} \max_{k=1}^{Tm} w(k/m) \frac{\hat{\sigma}_h}{\sigma_h} \hat{E}_{m,h}(k) \\
&= \max_{h=1}^{d} \max_{k=1}^{Tm} w(k/m) \max_{j=0}^{k-1} \frac{k-j}{\sqrt{m}\sigma_h} \left| \hat{\mu}_{m+j+1}^{m+k}(h) - \hat{\mu}_1^{m+j}(h) \right|
\end{aligned}$$

a version of the statistic $\widehat{\mathcal{T}}_{m,d}$, where all component-wise long-run variance estimators $\hat{\sigma}_h$ have been replaced by the (unknown) true long-run variances σ_h.

In the remainder of this paper, we will work with the sequences a_d, b_d defined in (3.6) with $q := q(T) = T/(T+1)$. The following theorem yields the asymptotic distribution of $\mathcal{T}_{m,d}$ as $m, d \to \infty$.

Theorem 3.6 *Suppose that the null hypothesis H_0 defined in (2.2) holds. Under the Assumptions 3.1- 3.4 it follows that*

$$a_d \big(\mathcal{T}_{m,d} - b_d \big) \overset{\mathcal{D}}{\Longrightarrow} G \,, \qquad \text{as} \ m, d \to \infty \,,$$

where G denotes a standard Gumbel random variable.

Note that due to the choice of a_d, b_d the limit distribution does not depend on the monitoring parameter T, which controls the length of the monitoring period, while the statistic $\widehat{\mathcal{T}}_{m,d}$ does depend on T.

Given Theorem 3.6 our final task is to identify suitable long-run variance estimators to obtain the asymptotic distribution of $\widehat{\mathcal{T}}_{m,d}$. We will identify a

general condition on the estimators in Assumption 3.7, which guarantees that all true long-run variances $\{\sigma_h^2\}_{h=1,\ldots,d}$ in the statistic $\mathcal{T}_{m,d}$ can be replaced by their corresponding estimators. Explicit estimators satisfying this assumption are constructed in Remark 3.8.

Assumption 3.7 Suppose that there exists a long-run variance estimator $\hat{\sigma}_h = \hat{\sigma}_h(m)$ based only on the stable initial set, such that $\mathbb{P}\big(\max_{h=1}^d |\hat{\sigma}_h - \sigma_h| \geq m^{-\delta_\sigma}\big) \leq c'_\sigma m^{-C'_\sigma}$, where c'_σ is a sufficiently large constant and $C'_\sigma > 0$, $\delta_\sigma > 0$ are sufficiently small constants.

Remark 3.8. In the field of sequential change point detection it is common to use only the initial stable data for the estimation of the long-run variance as this ensures that the estimate cannot be corrupted by a change [see for instance Aue et al. (2012), Wied and Galeano (2013) or Fremdt (2014) among many others]. It follows from Jirak (2015b) that Assumption 3.7 holds for the standard long-run variance estimators

$$\hat{\sigma}_{h,strd}^2 = \hat{\phi}_{0,h} + 2\sum_{t=1}^{H_m} \hat{\phi}_{t,h}\,, \qquad h = 1, \ldots, d\,, \tag{3.7}$$

where $\hat{\phi}_{t,h}$ denotes the lag t auto-covariance estimator in component h, that is

$$\hat{\phi}_{t,h} := \frac{1}{m-t} \sum_{i=t+1}^m \big(X_{i,h} - \hat{\mu}_1^m(h)\big)\big(X_{i-t,h} - \hat{\mu}_1^m(h)\big)\,. \tag{3.8}$$

The bandwidth parameter H_m in (3.7) is bounded by m^η for some $\eta = \eta(p, D)$ that fulfills constraints with respect to the constants D and p from Assumptions 3.1 and 3.3 [see Assumption 2.2 in Jirak (2015b)].

Corollary 3.9 *If the null hypothesis H_0 defined in* (2.2) *holds and the Assumptions* 3.1- 3.4 *and* 3.7 *are satisfied, it follows that*

$$a_d\big(\widehat{\mathcal{T}}_{m,d} - b_d\big) \overset{\mathcal{D}}{\Longrightarrow} G\,, \qquad as\ m, d \to \infty\ .$$

where G denotes a standard Gumbel random variable. In particular, $\widehat{\mathcal{T}}_{m,d}/\sqrt{\log d}$ converges in probability to $\sqrt{2q(T)}$.

If $g_{1-\alpha}$ denotes the $(1-\alpha)$ quantile of the standard Gumbel distribution, we obtain from Corollary 3.9 that the sequential procedure defined by (2.7) with $c_{d,\alpha} := g_{1-\alpha}/a_d + b_d$ has asymptotic size α, i.e.

$$\lim_{m,d\to\infty} \mathbb{P}_{H_0}\Big(\widehat{\mathcal{T}}_{m,d} > \frac{g_{1-\alpha}}{a_d} + b_d\Big) = \alpha\,. \tag{3.9}$$

The next theorem yields consistency of this monitoring scheme under the alternative hypothesis of a change in the mean vector.

Theorem 3.10 *Under the alternative hypothesis H_1 defined in (2.2), assume that there is a component h^* and a time point $k^* = k^*(m)$ such that*

$$\sqrt{\frac{m}{\log m}} \cdot \left| \mu_{m+k^*-1,h^*} - \mu_{m+k^*,h^*} \right| \to \infty \quad and \quad \limsup_{m \to \infty} \frac{k^*}{m} < T \ . \qquad (3.10)$$

If Assumptions 3.1- 3.4 and 3.7 are satisfied, it follows that

$$\lim_{m,d \to \infty} \mathbb{P}_{H_1}\left(\widehat{\mathcal{T}}_{m,d} > \frac{g_{1-\alpha}}{a_d} + b_d \right) = 1 \ .$$

Condition (3.10) shows that the test is able to detect alternatives which converge to the null hypothesis at the rate of $m^{-1/2}$ up to a factor $c_m\sqrt{\log m}$ with a sequence $(c_m)_{m \in \mathbb{N}}$ tending to ∞ at an arbitrary slow rate. This factor is needed to address for the high dimensional setting. In view of Assumption 3.1, the factor $\sqrt{m/\log m}$ may be replaced with $\sqrt{m/\log d}$ in (3.10), since they of the same order. Note also the time $m + k^*$ of the change is not permitted to be close to the end $m + mT$, which reflects the necessity to have a reasonably large sample after the change point such that the corresponding means can be estimated with sufficient precision.

Remark 3.11. In this paper, we used the maximum statistic for the aggregation of the different components, which has been used before for other problems [see, for example, Tartakovsky et al. (2006)]. Alternative statistics for aggregation could be used as well and there are numerous possibilities to choose from. The choice of the statistic depends on the type of alternative, which one is interested in. For example, robust schemes as proposed by Mei (2010) are based on the sum of (weighted) local statistics. In the present context, such a statistic reads as

$$\sum_{h=1}^{d} \max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h}(k) \ ,$$

which we suspect to be asymptotic Gaussian under appropriate standardization. It has been argued (mainly by simulations) that maximum-type statistics are more effective than sums when potential changes occur in only a few data streams [see Mei (2010); Xie and Siegmund (2013); Zou et al. (2015)]. On the other hand, when the change occurs in a moderate or large number of components, sum-type statistics outperform the maximum. We also refer to the work Fan et al. (2015) and Kock and Preinerstorfer (2019) for techniques to boost the power of testing high-dimensional parameters.

An improvement could also be obtained using the spatial information to construct a detector from the individual components. In the context of offline change point detection such a strategy was proposed by Wang and Samworth (2018) for the case of independent Gaussian data (in time and space), where sample splitting is used to estimate "optimal directions". They also indicate how the methodology can be extended to either temporally or spatially independent Gaussian data. In principle, such a strategy could be used in online monitoring of

(non-Gaussian) time and stationary dependent data with a stable initial sample. Alternatively, one could partition the estimate of the covariance matrix from the first step into blocks of highly correlated elements, aggregate the components corresponding to each block by a sum and finally calculate the maximum of sums corresponding to the different blocks. As indicated by the discussion in Wang and Samworth (2018) a theoretical analysis of such a two stage procedure in the general case considered in this paper is very difficult (maybe intractable). Moreover, independently of the theoretical aspects, some care is necessary, if such strategies are used in applications. In particular, it requires estimates of the covariance structure with sufficient precision. If the estimates from the first step have a too large variability adapted aggregation schemes may perform worse than non-adapted. To our knowledge, this phenomenon has not been discussed in the context of aggregation of statistics from individual components of high dimensional time series but it is well known in other circumstances such as adaptive designs – see, for example, Dette et al. (2013). For high dimensional time series it is difficult to obtain estimates of the (long-run) covariances with sufficient precision such that it can be used in a two-step procedure. Therefore we do not recommend the use of an adaptive aggregation in this context.

3.2. Bootstrap quantiles

Equation (3.9) and Theorem 3.10 show that the new sequential testing procedure (2.7) with $c_{d,\alpha} = g_{1-\alpha}/a_d + b_d$ has asymptotic level α and is consistent. However, the approach so far is based on an approximation of the distribution of the statistic by a Gumbel distribution featuring the well-known disadvantage that the convergence rates in such limiting results are rather slow. As a consequence for small sample sizes, these quantiles may yield slightly imprecise approximations in practical applications. To tackle this problem, we will propose a bootstrap procedure. Note that the development of resampling procedures in the sequential regime is a difficult problem. On the one hand, critical values can be computed only from the initial stable sample, but this set can be too small to obtain reliable values. On the other hand, one can compute new critical values with each new data point, which is computationally expensive and can be corrupted by an undetected structural break. Therefore, both approaches have natural advantages and disadvantages. For i.i.d. data Kirch (2008) proposes a bootstrap procedure for sequential detection of a structural break in the mean of an one-dimensional sequence by a combination of both methods following ideas of Steland (2006). Nevertheless, the construction of bootstrap methodology for sequential change point detection in the high dimensional regime remains challenging.

To be precise, note that by Lemma A.2 and A.3 used in the proof of Theorem 3.6 in the online supplement we obtain the approximation

$$\mathbb{P}\Big(a_d\big(\mathcal{T}_{m,d} - b_d\big) \leq x\Big) - \mathbb{P}\Big(a_d\big(\mathcal{T}_{m,d}^{(Z)} - b_d\big) \leq x\Big) = o(1) \; . \qquad (3.11)$$

Here the statistic $\mathcal{T}_{m,d}^{(Z)}$ is the counterpart of $\mathcal{T}_{m,d}$ computed from standard Gaussian random variables $Z_{t,h}$, which are independent in time and have spatial dependence structure $\mathrm{Cov}\left(Z_{0,h}, Z_{0,i}\right) = \rho_{h,i}$, where $\rho_{h,i}$ are the long-run correlations defined in (3.4). In view of the approximation (3.11), it is therefore reasonable to obtain the quantiles for the statistic $\mathcal{T}_{m,d}$ from those of the statistic $\mathcal{T}_{m,d}^{(Z)}$, which can easily be simulated if the correlations $\rho_{h,i}$ were known. These parameters can be straightforwardly estimated from the initial stable data set $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$.

Compared to a bootstrap procedure continuously performed during monitoring, this idea exhibits two important advantages. Firstly, it ensures that the correlation estimates cannot be corrupted by a mean change, that may occur during the monitoring period. Secondly, it requires less computational effort, as the quantile is only computed once before monitoring is commenced. This is of vital importance in a high dimensional setup, where the method on its own is already quite expensive and resampling and/or repeated estimations during monitoring may quickly exceed the computational resources.

Before discussing the technical details of this resampling procedure, we state a necessary assumption regarding the precision of the estimates of the long-run covariances.

Assumption 3.12 Suppose that there exists a long-run covariance estimator $\hat{\gamma}_{h,i} = \hat{\gamma}_{h,i}(m)$ based on the stable initial data set $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$, such that $\mathbb{P}\left(\max_{h,i=1}^{d} \left|\hat{\gamma}_{h,i} - \gamma_{h,i}\right| \geq m^{-\delta_\gamma}\right) \leq c_\gamma m^{-C_\gamma}$, where c_γ is a sufficiently large constant and $C_\gamma > 0$, $\delta_\gamma > 0$ are sufficiently small constants.

Remark 3.13. A canonical choice for a long-run covariance estimator that satisfies Assumption 3.12, is the standard estimator

$$\hat{\gamma}_{h,i,strd} = \hat{\phi}_{0,h,i} + \sum_{t=1}^{H_m} \hat{\phi}_{t,h,i} + \sum_{t=1}^{H_m} \hat{\phi}_{t,i,h} \ ,$$

where H_m is an appropriate bandwidth and the involved cross-components covariance estimators are given by $\hat{\phi}_{t,h,i} := (m-t)^{-1} \sum_{j=t+1}^{m} \left(X_{j,h} - \hat{\mu}_1^m(h)\right) \left(X_{j-t,i} - \hat{\mu}_1^m(i)\right)$. Note that these definitions are natural extensions of the long-run variance and auto-covariance estimators in (3.7) and (3.8), respectively. Therefore one can use similar arguments as given in Jirak (2015b) (for the verification of Assumption 3.7) to prove the consistency stated in Assumption 3.12.

In the following, denote by $\mathcal{X} = \sigma(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m)$ the σ-algebra generated by the initial sample and let $\mathbb{P}_{|\mathcal{X}}, \mathrm{Cov}_{|\mathcal{X}}$ denote the conditional probability and covariance with respect to $\mathcal{X}$. To define the bootstrap statistic, let

$$\left\{ \widehat{\boldsymbol{Z}}_t = \left(\widehat{Z}_{t,1}, \ldots, \widehat{Z}_{t,d}\right)^{\top} \right\}_{t=1,\ldots,m+Tm} \tag{3.12}$$

denote centered random vectors, that are – conditionally on $\mathcal{X}$ – independent and Gaussian distributed with covariance structure

$$\mathrm{Cov}_{|\mathcal{X}} \left(\widehat{Z}_{t,h}, \ \widehat{Z}_{t,i}\right) = \hat{\rho}_{h,i} \ , \tag{3.13}$$

where $\hat{\rho}_{h,i}$ are correlation estimators canonically defined by $\hat{\rho}_{h,i} = \hat{\gamma}_{h,i}/(\hat{\sigma}_h \hat{\sigma}_i)$ for $h \neq i$ and

$\hat{\rho}_{h,h} = 1$. Note that by the definition in (3.13) the random vectors preserve the (estimated) spatial correlation structure of the time series. Next, denote the component-wise mean estimators for subsamples of $\{\widehat{Z}_{t,h}\}_{t=1}^{m+Tm}$ by $\hat{z}_i^j(h) := (j - i + 1)^{-1} \sum_{t=i}^{j} \widehat{Z}_{t,h}$ and the final bootstrap statistic by

$$\widehat{\mathcal{T}}_{m,d}^{(Z)} := \max_{k=1}^{Tm} \max_{h=1}^{d} \max_{j=0}^{k-1} \frac{w(k/m)(k-j)}{\sqrt{m}} \left| \hat{z}_{m+j+1}^{m+k}(h) - \hat{z}_1^{m+j}(h) \right| . \qquad (3.14)$$

Once the correlation estimates $\hat{\rho}_{h,i}$ are computed, the conditional distribution of $\widehat{\mathcal{T}}_{m,d}^{(Z)}$ can be approximated by Monte-Carlo simulations with arbitrary precision, generating replicates of $\{\widehat{Z}_{t,h}\}_{t=1,\ldots,m+mT}^{h=1,\ldots d}$. Thus, provided with a batch of realizations of the statistic $\widehat{\mathcal{T}}_{m,d}^{(Z)}$, one can compute the corresponding empirical quantile for the desired test level and launch the sequential procedure with this bootstrap quantile instead of the (possibly less precise) Gumbel quantile. The following result yields the validity of our proposed bootstrap procedure.

Theorem 3.14 (Bootstrap consistency) *Under the Assumptions 3.1–3.4 and 3.12 we have*

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}_{|\mathcal{X}} \left(a_d \big(\widehat{\mathcal{T}}_{m,d}^{(Z)} - b_d \big) \leq x \right) - \mathbb{P}_{H_0} \left(a_d \big(\widehat{\mathcal{T}}_{m,d} - b_d \big) \leq x \right) \right| = o_{\mathbb{P}}(1) \ ,$$

where $\mathbb{P}_{H_0}$ denotes the probability under the null hypothesis of no change in any component.

Combining Theorem 3.14 with Corollary 3.9 and Theorem 3.10, it follows that the use of the quantiles of the bootstrap distribution in (2.7) yields a consistent monitoring scheme, which keeps its pre-specified nominal level.

Remark 3.15. Note that the bootstrap procedure uses correlation estimates from the stable sample, but does not use these estimates for the aggregation of the componentwise detectors in the monitoring scheme.

We conclude this section by presenting a detailed algorithm to monitor for non-simultaneous change points in the mean vector of a high dimensional time series at test level α. For this purpose, we denote the set of all components without a change in the mean by

$$\mathcal{S}_d := \big\{ h \in \{1, \ldots, d\} \mid \mu_{1,h} = \mu_{2,h} = \cdots = \mu_{m+Tm,h} \big\} \ .$$

For the complement set $\mathcal{S}_d^c = \{1, \ldots, d\} \setminus \mathcal{S}_d$ let $k_h^* \in \{1, \ldots, mT\}$ denote the time of change in the component $h \in \mathcal{S}_d^c$ within the monitoring period, that is

$$\mu_{1,h} = \cdots = \mu_{m,h} = \cdots = \mu_{m+k_h^*-1,h} \neq \mu_{m+k_h^*,h} = \cdots = \mu_{m+Tm,h} \ .$$

Note that, in this model, there is still at most one change point per component, however we allow the change times to be different among components. For what

follows, recall the construction of the bootstrap statistic from Section 3.2. The algorithm below is capable of identifying the sets $\mathcal{S}_d$ and $\mathcal{S}_d^c$, see Theorem 3.17 below.

Algorithm 3.16.

Step 1: Either choose the quantile $q_{1-\alpha}$ using the approximation by the Gumbel distribution, that is $q_{1-\alpha} = g_{1-\alpha}$, where $g_{1-\alpha}$ is the $(1-\alpha)$-quantile of the Gumbel distribution or, alternatively, obtain the quantile from the bootstrap as follows:

Step 1.1: Compute the long-run correlation estimates $(\hat{\rho}_{i,j})_{i,j=1}^{d}$ from the initial set $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$.

Step 1.2: Based on these estimates, generate N independent realizations of the Gaussian vectors $(\widehat{Z}_{t,1}, \ldots, \widehat{Z}_{t,d})^{\top}$ with covariance structure $(\hat{\rho}_{i,j})_{i,j=1}^{d}$ for $t = 1, \ldots, m + Tm$ and compute the corresponding bootstrap statistics

$$a_d\big(\widehat{\mathcal{T}}_{m,d}^{(Z)}(1) - b_d\big), a_d\big(\widehat{\mathcal{T}}_{m,d}^{(Z)}(2) - b_d\big), \ldots, a_d\big(\widehat{\mathcal{T}}_{m,d}^{(Z)}(N) - b_d\big)$$

defined in (3.14).

Step 1.3: Compute $q_{1-\alpha}$ as the empirical $(1-\alpha)$-quantile of the sample

$$\left\{ a_d\left(\widehat{\mathcal{T}}_{m,d}^{(Z)}(n) - b_d\right) \right\}_{n=1,\ldots,N} .$$

Step 2: Initialize $\widehat{\mathcal{S}}_{d,\alpha} := \{1, \ldots, d\}$ and set $k = 1$.

Monitoring: While $k \leq Tm$ compute the statistics $\hat{E}_{m,h}(k)$. If the inequality

$$\max_{h \in \widehat{\mathcal{S}}_{d,\alpha}} w(k/m)\hat{E}_{m,h}(k) > \frac{q_{1-\alpha}}{a_d} + b_d$$

holds, reject the null hypothesis in favor of the alternative. Eliminate the components that led to the rejection, i.e.

$$\widehat{\mathcal{S}}_{d,\alpha} \longleftarrow \widehat{\mathcal{S}}_{d,\alpha} \setminus \left\{ h \in \widehat{\mathcal{S}}_{d,\alpha} \mid w(k/m)\hat{E}_{m,h}(k) > \frac{q_{1-\alpha}}{a_d} + b_d \right\}$$

and continue monitoring for $k \longleftarrow k+1$ with the remaining components in $\widehat{\mathcal{S}}_{d,\alpha}$.

Step 3: If there was no rejection during monitoring, decide for the null hypothesis of no change in the mean vector. In case of rejections, decide for the alternative of a change in at least one component. Then, it holds that

$$\widehat{\mathcal{S}}_{d,\alpha} = \left\{ h \in \{1, \ldots, d\} \mid \max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h}(k) \leq \frac{q_{1-\alpha}}{a_d} + b_d \right\} \quad (3.15)$$

and the components remaining in this set are assumed as mean stable.

The following theorem states, that Algorithm 3.16 is able to separate the sets $\mathcal{S}_d$ and $\mathcal{S}_d^c$ correctly.

Theorem 3.17 *Let Assumptions 3.1-3.4 and 3.12 be satisfied and assume that Algorithm 3.16 was launched either with the Gumbel or with the bootstrap quantile. The set $\widehat{\mathcal{S}}_{d,\alpha}$ defined in (3.15) satisfies*

$$\limsup_{m,d\to\infty} \mathbb{P}\Big(\mathcal{S}_d \subset \widehat{\mathcal{S}}_{d,\alpha}\Big) \geq 1 - \alpha \ . \tag{3.16}$$

If further

$$\sqrt{\frac{m}{\log m}} \ \min_{h\in\mathcal{S}_d^c} |\mu_{m+k_h^*-1,h} - \mu_{m+k_h^*,h}| \to \infty \ \ and \ \ \limsup_{m,d\to\infty} \max_{h\in\mathcal{S}_d^c} \frac{k_h^*}{m} < T, \tag{3.17}$$

then

$$\lim_{m,d\to\infty} \mathbb{P}\Big(\mathcal{S}_d \supset \widehat{\mathcal{S}}_{d,\alpha}\Big) = 1 \ . \tag{3.18}$$

By Theorem 3.17, the set $\widehat{\mathcal{S}}_{d,\alpha}$ contains all the components without a change in the mean with high probability. It may also contain components in which the change in the mean is so small that it cannot be reliably detected. However, if we assume that that all changes in the mean are large enough and not too close to the monitoring end [see (3.17)], then $\widehat{\mathcal{S}}_{d,\alpha} = \mathcal{S}_d$ with high probability. That is, one detects all components exhibiting a change point without causing a false alarm.

4. Finite sample properties

In this section, we investigate the finite sample properties of the new monitoring schemes by means of a simulation study and illustrate potential applications in a data example.

4.1. Simulation study

In our simulation study we consider the following models:

(M1) $X_{t,h} = \theta_{t,h}$,   (M2) $X_{t,h} = 0.1X_{t-1,h} + \varepsilon_{t,h}$,

(M3) $X_{t,h} = \eta_{t,h} + 0.3\eta_{t-1,h} - 0.1\eta_{t-2,h}$,   (M4) $X_{t,h} = \tilde{\varepsilon}_{t,h}$,

where $\{\theta_{t,h}\}_{t\in\mathbb{N},h\in\mathbb{N}}$ is an array of i.i.d. random variables that follow a $\chi_{10}^2/\sqrt{20}$-distribution, $\{\varepsilon_{t,h}\}_{t\in\mathbb{N},h\in\mathbb{N}}$ is an array of i.i.d. standard Gaussian random variables, $\{\eta_{t,h}\}_{t\in\mathbb{N},h\in\mathbb{N}}$ is an array of i.i.d. Laplace(0,1) distributed random variables and $\{\tilde{\varepsilon}_{t,h}\}_{t\in\mathbb{N},h\in\mathbb{N}}$ are random variables, such that the d-dimensional vectors $\big\{\tilde{\boldsymbol{\varepsilon}}_t = (\tilde{\varepsilon}_{t,1},\ldots,\tilde{\varepsilon}_{t,d})^\top\big\}_{t\in\mathbb{Z}}$ are i.i.d., centered Gaussian random vectors with covariance structure $\text{Cov}(\tilde{\varepsilon}_{1,j}, \tilde{\varepsilon}_{1,i}) = 1/(|j-i|+1)$. For the alternative hypothesis,

we also consider the models (M1)-(M4) and add a shift in the mean, at some point $m + k^*$, that is

$$X_{t,h}^{(\delta,A)} = \begin{cases} X_{t,h} & \text{if } t < m + k^* , \\ X_{t,h} + \delta \cdot I\{h \in A\} & \text{if } t \geq m + k^* , \end{cases} \tag{4.1}$$

where I denotes the indicator function, A is the set of spatial components affected by the change and δ is the size of the change. In order to examine the influence of both parameters on the procedure, we will consider different values of δ and three different choices of the set A below.

For the long-run variance estimation, we use the quadratic spectral kernel estimator [see Andrews (1991)] in each component, that is

$$\hat{\sigma}_h^2 := \sum_{|t| \leq m-1} k\left(\frac{t}{H_m}\right)\hat{\phi}_{t,h} , \tag{4.2}$$

with the empirical auto-covariances $\hat{\phi}_{t,h} := m^{-1} \sum_{i=t+1}^m \left(X_{i,h} - \hat{\mu}_1^m(h)\right)\left(X_{i-t,h} - \hat{\mu}_1^m(h)\right)$ and underlying kernel

$$k(x) := \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right) .$$

In particular, we employ the implementation of the estimator (4.2) provided by the R-package 'sandwich' [see Zeileis (2004)] and select the bandwidth parameter as $H_m = \log_{10}(m)$. Note that we only use the stable set $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ for the estimation of the long-run variance, which avoids corruption from observations after the potential change point under the alternative [see the discussion in Remark 3.8]. All results presented in this section are based on 1000 simulation runs and for the AR(1)-process (M2) we employ a burn-in sample of 200 observations. The test level is always fixed at $\alpha = 0.05$.

In Table 1 and 2, we illustrate the finite sample properties of the detection scheme (2.7) under the null hypothesis for different choices of the sample size m, the dimension d and the length of the monitoring period determined by T. The results in Table 1 are based on the weak convergence in Corollary 3.9 and therefore we use the critical value $c_{d,\alpha} = g_{1-\alpha}/a_d + b_d$ in (2.7), where $g_{1-\alpha}$ is the quantile of the Gumbel distribution.

The results in Table 2 are obtained by the bootstrap procedure as described in Step 2 of Algorithm 3.16 in the online supplement. We note that the use of (any) sequential monitoring scheme in a simulation study is computationally demanding, in particular for a high dimensional setup. In our case, we have to estimate the spatial correlation structure in each simulation run and then simulate the quantile of the distribution of the statistic $\widehat{\mathcal{T}}_{m,d}^{(Z)}$ defined in (3.14). Of course, this is no problem in data analysis as in this case the monitoring procedure has only to be conducted once, but it requires large computational resources in a simulation, where the same procedure is repeated 1000 times. Therefore, in order to reduce the computational complexity of the bootstrap approach in

TABLE 1

Approximation of the nominal level by the detector defined in (2.7) *for different choices of initial sample size m, dimension d and monitoring duration $m \cdot T$. Critical values are obtained from Corollary* 3.9 *(approximation by Gumbel distribution). The nominal level is $\alpha = 0.05$.*

T	model	m=100		m=200		m=500	
		d=100	d=200	d=200	d=500	d=200	d=500
1	(M1)	10.4%	13.7%	6.7%	10.2%	4.5%	6.1%
	(M2)	7.4%	11.3%	8.0%	9.4%	4.1%	4.9%
	(M3)	5.8%	7.6%	4.2%	7.4%	3.6%	3.2%
	(M4)	4.2%	7.3%	4.1%	5.6%	2.8%	3.8%
2	(M1)	9.9%	12.9%	7.7%	10.5%	3.8%	6.6%
	(M2)	9.0%	11.2%	7.5%	9.2%	5.0%	4.4%
	(M3)	5.7%	7.4%	4.6%	5.1%	2.3%	2.8%
	(M4)	4.7%	7.3%	5.2%	7.0%	3.0%	3.7%
4	(M1)	10.9%	16.6%	8.9%	11.1%	5.0%	5.5%
	(M2)	9.1%	12.6%	7.7%	9.5%	3.8%	5.4%
	(M3)	6.0%	8.4%	4.6%	5.6%	2.4%	2.9%
	(M4)	6.3%	7.2%	5.2%	5.8%	2.7%	5.5%

TABLE 2

Approximation of the nominal level using the detector defined in (2.7) *for different choices of initial sample size m and dimension d. Critical values are computed by the bootstrap with spatial independence. The nominal level is $\alpha = 0.05$.*

T	model	m=100		m=200		m=500	
		d=100	d=200	d=200	d=500	d=200	d=500
1	(M1)	9.7%	13.1%	7.3%	8.5%	9.8%	8.4%
	(M2)	6.9%	11.1%	8.1%	7.0%	8.1%	6.8%
	(M3)	5.2%	7.5%	4.2%	5.6%	6.4%	4.7%
	(M4)	4.0%	7.0%	4.1%	4.9%	6.4%	5.5%

the simulation study, we do not estimate the spatial correlation structure for the bootstrap but employ temporal and spatial independent Gaussian random variables to generate the bootstrap statistics defined in (3.14). With this adaption, the quantiles are fixed within each column of Table 2, which makes the simulation study practicable. Moreover, it can easily be seen from the theory developed in Section 3 that the use of these quantiles also yields a consistent test in (2.7). Here we point to Lemma A.1 in Appendix A in the online supplement.

In Table 1, we observe a reasonable approximation of the nominal level by the asymptotic test in many cases, which becomes more accurate with larger initial sample size m and dimension d. For instance, consider the model (M2) for the choice $T = 2$, where we have obtained a type I error of 9.0% for $m = d = 100$ and 7.5% for $m = d = 200$. This finally reduces to an appropriate approximation of 4.4% for the choice $m = d = 500$. As common for high dimensional procedures, the relation of sample size m and dimension d has a severe impact on the performance of the monitoring procedure. For example, an empirical type I error of 4.2% was measured for model (M4) with $m = d = 100$ and $T = 1$,

which increases to 7.3% if the dimension is set to $d = 200$. This effect becomes weaker, when the sample size is generally increased.

In Table 2, we display the type I error for the method where the quantiles are calculated by the bootstrap as described above. For the sake of brevity, we focus on the case $T = 1$, as the results obtained for different choices of T in Table 1 are similar. We observe a very reasonable approximation of the desired test level and – compared to the results in Table 1 – some improvement by the bootstrap procedure for small sample sizes. In particular, application of the bootstrap is recommended in cases where m and d are relatively small, where the approximation of the nominal level using the quantiles from the Gumbel distribution is rather imprecise, while the extra computational costs for the bootstrap are still tolerable.

To analyze the performance of the sequential procedure under the alternative hypothesis we consider the model (4.1), where the processes $X_{t,h}$ are defined by (M1)-(M4). We distinguish between the following three scenarios:

(A1) The change occurs only in one component. This corresponds to the choice
      $A = \{1\}$.
(A2) The change occurs in 50% of the components, i.e. $A = \{1, \ldots, d/2\}$.
(A3) The change occurs in all components, i.e. $A = \{1, \ldots, d\}$.

For the sake of brevity and readability, we focus on the case $T = 1$ under the alternative and only consider change positions in the middle of the monitoring period, i.e. we fix $k^* = m/2$. In Figures 1 and 2, we display the rejection probabilities of the detection rule (2.7) for these scenarios, different values of the change, different sample sizes and dimensions. The critical values in (2.7) are given by $c_{d,\alpha} = q/a_d + b_d$, where q is the quantile obtained by the spatial independent bootstrap, which appears more accurate than the quantiles derived from the Gumbel distribution.

The results can be summarized as follows. In all considered scenarios the new monitoring procedure (2.7) for a change in the high dimensional mean vector has reasonable power under the alternative, and in all cases the type II error approaches zero for an increasing size δ of the change. As expected, the power is lower under alternative (A1), where the change occurs in only one coordinate. To give an example, consider model (M1) and (M2) corresponding to the left columns in Figures 1 and 2. The results for the different alternatives (A1), (A2) and (A3) can be found in the first, second and third rows of the figures, respectively. If the sample size and dimension are given by $m = 100$ and $d = 100$ we observe from Figure 1 that for $\delta = 0.7$ the power for model (M1) and (M2) under alternative (A1) is approximately given by 0.25, while it is close to one under alternative (A2). Interestingly, the differences between alternatives (A2) and (A3) are not so strong, but they are still clearly visible. For instance, a comparison between the left parts of the second and third row of Figure 2 shows that the power of the detection scheme (2.7) in model (M1) for $\delta = 0.3$ is approximately 0.88 for alternative (A2) and 0.99 for alternative (A3).
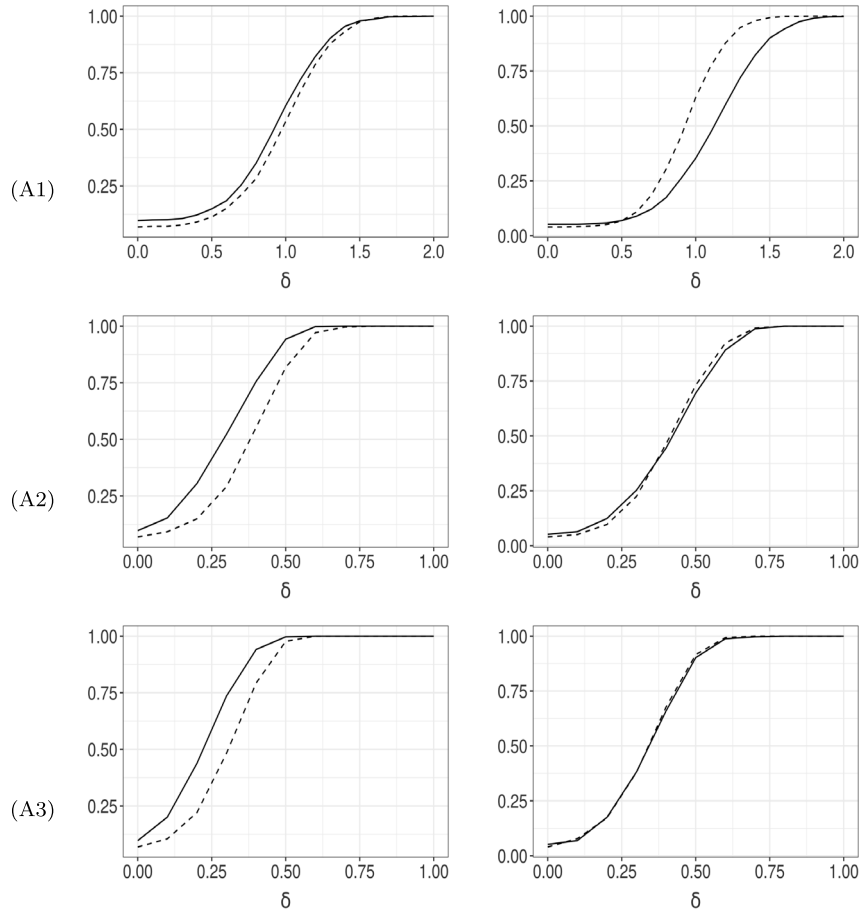
FIG 1. *Simulated power of the monitoring scheme* (2.7) *for different size δ of the change, sample size m = 100 and dimension d = 100. Left panels: Solid line (M1), dashed line (M2). Right panels: Solid line (M3), dashed line (M4). The nominal level is α = 0.05.*

The differences between the four data generating models are in general not substantial with one exception. In model (M3) under alternative (A1) the power of the detection scheme (2.7) is considerably smaller [see the first rows in Figure 1 and Figure 2].

We summarize the discussion of the finite sample properties emphasizing that our numerical results have supported the theoretical findings developed in Section 3 in all cases under consideration.

4.2. Data example

In this section, we illustrate potential applications of the new monitoring scheme in a data example. For this purpose, we consider a data set sampled in hydrol-
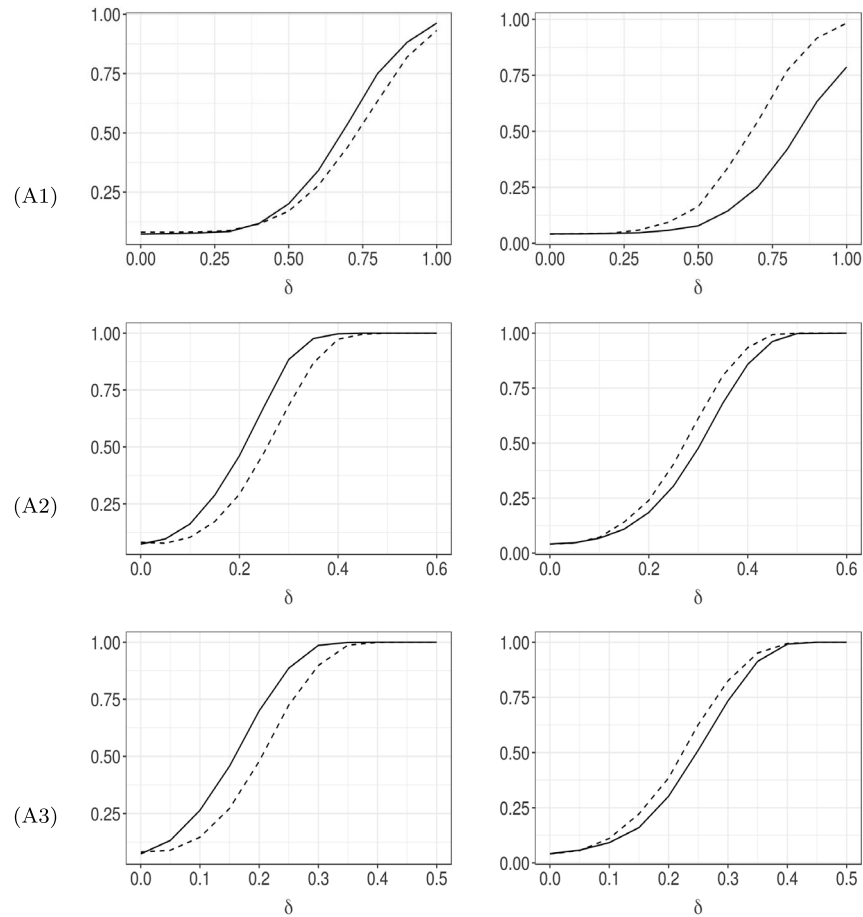
FIG 2. *Simulated power of the monitoring scheme* (2.7) *for different size* δ *of the change, sample size* $m = 200$ *and dimension* $d = 200$. *Left panels: Solid line (M1), dashed line (M2) Right panels: Solid line (M3), dashed line (M4). The nominal level is* $\alpha = 0.05$.

ogy, which consists of the average daily flows measured in m^3/sec of the river Chemnitz at Göritzhain in Saxony, Germany, for the years 1909-2013.

This data has been previously analyzed in the (retrospective) change point literature by Sharipov et al. (2016), who developed methodology for detecting change points in functional data. The data set consists of a sample of $n = 105$ observations with dimension $d = 365$, such that each vector $\boldsymbol{X}_t = (X_{t,1}, \ldots, X_{t,365})^\top$ contains the daily average flows of one (German) hydrological year, which lasts from 1st of November to 31st of October. For instance, the data point $X_{1,1}$ represents the daily average flow of the 1st of November 1909, while $X_{105,365}$ is the same key figure for the 31st of October 2014. By this transformation, Sharipov et al. (2016) located a change in the annual flow

TABLE 3

Structural breaks detected in the river flow data for a test level of $\alpha = 0.05$ (left column) and $\alpha = 0.01$ (right column). The column 'year' specifies the (hydrological) year after which the rejection occurred. For instance 1945 means that the data from the hyrodological year 01st of November 1945 to 31st of October 1946 was already under consideration.

$\alpha = 0.05$				$\alpha = 0.01$			
component	year	component	year	component	year	component	year
101	1945	54, 209	1995	101	1945	280	2009
252, 253	1953	264	1996	252, 253	1953	54, 215	2012
249, 251	1954	285, 286, 287	2001	249	1957	209	2013
247	1957	92	2002	105	1960		
105	1960	99,192	2003	189	1977		
189, 191	1977	138	2004	191	1979		
104	1979	280, 283	2009	100	1980		
100, 190	1980	44	2010	104	1986		
102	1986	55, 214, 215, 216	2012	53	1995		
53, 57	1993	199	2013	285, 286, 287	2001		

curves in the year 1964. Dette and Gösmann (2018) propose a retrospective test for relevant changes in a high dimensional time series. They consider the same data and locate 4 different mean changes that exceed a test threshold of 0.63 and are traced back to the dates 10th of July 1950, 18th of March 1956, 23rd of December 1965, 7th of February 1979, which correspond to spatial components 252, 137, 53, 99, respectively.

Based on these prior analyses, we consider the first 35 observations as our initial stable data set and will use the remaining 70 observations as the monitoring period corresponding to a choice of $m = 35$ and $T = 2$. From the initial set $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{35}$ the spatial correlation structure is estimated via the implementation of the estimator based on the quadratic spectral kernel provided in the R-package 'sandwich' [see Zeileis (2004)]. By the bootstrap in Algorithm 3.16 the quantiles are obtained as $q_{0.99} = 7.43$ and $q_{0.95} = 4.93$, for which we conduct our monitoring method. During the monitoring period, we proceed as precisely described in Algorithm 3.16: If the detection scheme rejects the null hypothesis of no change at a certain time point, we report the corresponding component(s) as unstable and remove it/them from the sample. Afterwards, we continue monitoring with the remaining components until there is another rejection or the end of the monitoring period is attained.

The results of this procedure are displayed in Table 3 for the test levels $\alpha = 0.01$ and $\alpha = 0.05$ and can be summarized as follows. For a test level of 0.05 more unstable components (33) are identified than for 0.01 (17 components). Naturally, all breaks identified by the lower test level, are also detected by the other one, while the time of detection is sometimes earlier in the latter case. As the data exhibits (positive) spatial correlation, breaks partially occur in clusters, for example consider components 285, 286 and 287, for which both test levels detect changes or components 214, 215 and 216, for which changes are found at test level 0.05. Apparently, these *clusters* have to be considered as one change affecting several components, for instance a flood event or a new seasonal effect.

It is worth mentioning, that our findings match three out of four unstable
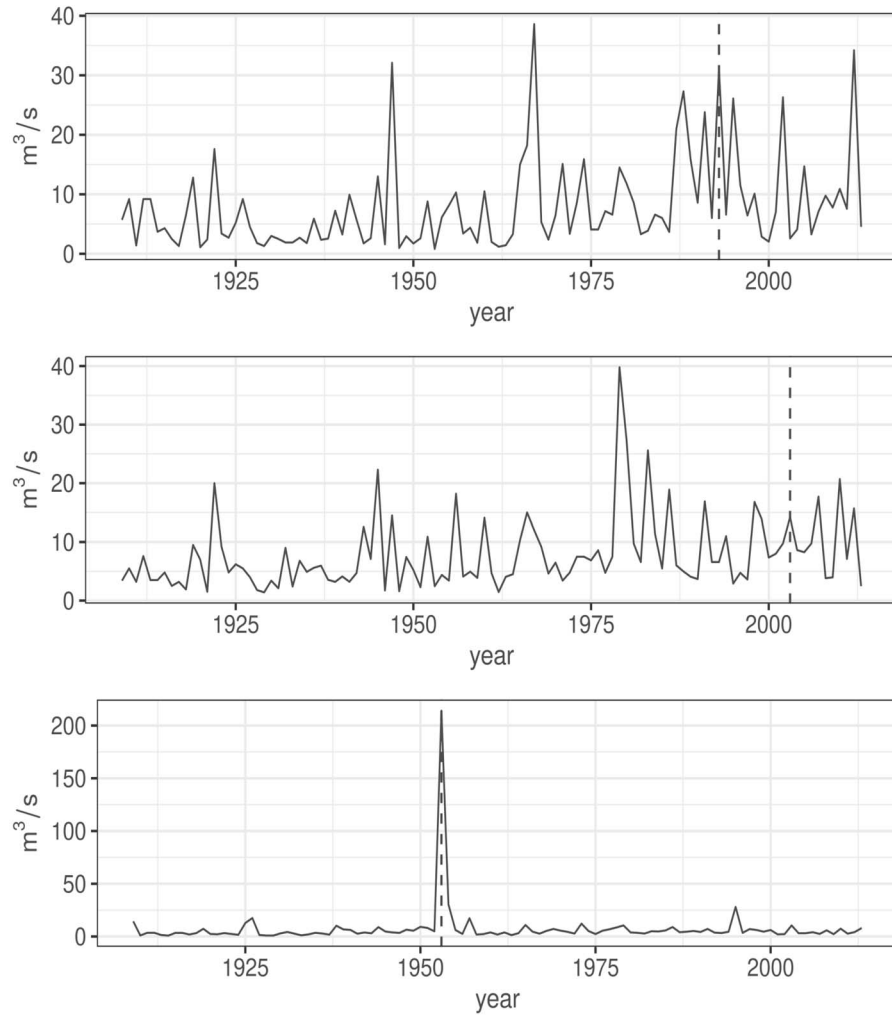
Fig 3. *Average daily flows for the dates 23rd of December (spatial component 53, upper row), 7th of February (spatial component 99, middle row) and 10th of July (spatial component 252, lower row). Vertical dashed lines indicate the time points at which a break was detected by the sequential method with a test level of 5%.*

components identified by the threshold procedure of Dette and Gösmann (2018). Namely, we refer to components 53, 99, 252, which are likewise identified to contain a break by our sequential analysis. To illustrate the data set, we finally display the average daily flow over the years for these three components in Figure 3. The plots indicate that the break in component 252 (10th of July) is most probably caused by a huge outlier in the hydrological year starting in November 1953, which leads to an immediate rejection. This observation can be easily linked to a flood event in Saxony in summer of 1954 [see for instance

Schröter et al. (2013)]. For the components 53 (23rd of December) and 99 (7th of February), a visual inspection of the plots indicates *actual* structural changes.

Appendix A: Proofs of main results

Let us first introduce some necessary notation, which will be frequently used throughout this section. The symbol $\lesssim$ denotes an inequality up to a constant, which does not depend on size m of the training sample and the dimension d. For two sequences e_n and f_n, we write $e_n \sim f_n$, whenever $f_n/e_n \to 1$ as $n \to \infty$.

Let $q \in (0,1]$. By $F_{\mathbb{M}}$ we denote the distribution function of the range of a (standard) Brownian motion, that is $\mathbb{M} = \max_{t \in [0,q]} W(t) - \min_{t \in [0,q]} W(t)$, which can be found in Borodin and Salminen (1996), page 146, and is given by

$$
F_{\mathbb{M}}(x) = \begin{cases} 1 + 4 \sum\limits_{k=1}^{\infty} (-1)^k k \operatorname{Erfc}\left(\dfrac{kx}{\sqrt{2q}}\right) & \text{if } x > 0, \\ 0 & \text{otherwise}, \end{cases}
$$

where $\operatorname{Erfc} = 1 - \operatorname{Erf}$ denotes the complementary error function. Using the elementary property $\operatorname{Erf}(x) = 2\Phi(x\sqrt{2}) - 1$ we obtain

$$
F_{\mathbb{M}}(x) = 1 + 8 \sum_{k=1}^{\infty} (-1)^k k \Phi\left(-\frac{kx}{\sqrt{q}}\right) \text{ for } x > 0, \tag{A.1}
$$

where throughout this paper Φ denotes the cumulative distribution function of a standard normal distribution.

A.1. Some preliminary results

We will begin with an auxiliary result. In Lemma A.1 we investigate the weak convergence of the maximum of independent identically distributed random variables with the same distribution as the random variable $\mathbb{M}$ defined in (2.6).

Lemma A.1 *Let $q \in (0,1]$ and $M_1', M_2', \ldots$ be independent identically distributed random variables with*

$$
M_1' \overset{\mathcal{D}}{=} \max_{0 \le t \le q} W(t) - \min_{0 \le t \le q} W(t),
$$

where W denotes a standard Brownian motion. Then, as $d \to \infty$, it holds that

$$
a_d \left(\max_{h=1}^{d} M_h' - b_d \right) \overset{\mathcal{D}}{\Longrightarrow} G,
$$

where G is a standard Gumbel distributed random variable and a_d and b_d are given by

$$
a_d = \sqrt{\frac{2 \log d}{q}} \quad \text{and} \quad b_d = \sqrt{2q \log d} - \frac{\sqrt{q}\,(\log \log d - \log \frac{16}{\pi})}{2\sqrt{2 \log d}}.
$$

Proof of Lemma A.1. The desired extreme value convergence will be derived from the distribution function of M_1', which is given by $F_{\mathbb{M}}$ in (A.1). Observe, that $F_{\mathbb{M}}$ is twice differentiable with derivatives (for $x > 0$)

$$F_{\mathbb{M}}'(x) = \frac{4\sqrt{2}}{\sqrt{\pi q}} \sum_{k=1}^{\infty} (-1)^{k+1} k^2 \exp\left(-\frac{k^2 x^2}{2q}\right),$$

$$F_{\mathbb{M}}''(x) = \frac{4\sqrt{2}}{\sqrt{\pi q} q} \sum_{k=1}^{\infty} (-1)^{k} k^4 x \exp\left(-\frac{k^2 x^2}{2q}\right),$$

where we used that the series converge uniformly on all intervals $[\varepsilon, \infty)$ for $\varepsilon > 0$ and therefore term by term differentiation is allowed. Thus, by Theorem 1.1.8 from de Haan and Ferreira (2006) the distribution function $F_{\mathbb{M}}$ is in the domain of attraction of the Gumbel distribution if

$$\lim_{x \to \infty} \frac{(1 - F_{\mathbb{M}}(x)) F_{\mathbb{M}}''(x)}{(F_{\mathbb{M}}'(x))^2} = -1. \qquad (A.2)$$

Next, we prove (A.2). To this end, note that by the definition of the complimentary error function, we obtain for $x > 0$

$$x\left(1 - F_{\mathbb{M}}(x)\right) \exp\left(\frac{x^2}{2q}\right) = 4 \sum_{k=1}^{\infty} (-1)^{k+1} kx \, \mathrm{Erfc}\left(\frac{kx}{\sqrt{2q}}\right) \exp\left(\frac{x^2}{2q}\right) \qquad (A.3)$$

$$= A_1(x) + A_2(x),$$

where the two summands on the right-hand side are given by

$$A_1(x) = \frac{8}{\sqrt{\pi}} \int_{\frac{x}{\sqrt{2q}}}^{\infty} \exp(-\tau^2) d\tau \cdot x \exp\left(\frac{x^2}{2q}\right),$$

$$A_2(x) = \frac{8}{\sqrt{\pi}} \sum_{k=2}^{\infty} (-1)^{k+1} kx \int_{\frac{kx}{\sqrt{2q}}}^{\infty} \exp(-\tau^2) d\tau \cdot \exp\left(\frac{x^2}{2q}\right).$$

We will treat the two summands of the last display separately. For the first summand we obtain that

$$\lim_{x \to \infty} A_1(x) = \frac{8}{\sqrt{\pi}} \lim_{x \to \infty} \frac{\int_{\frac{x}{\sqrt{2q}}}^{\infty} \exp(-\tau^2) d\tau}{x^{-1} \exp\left(\frac{-x^2}{2q}\right)}$$

$$= \frac{8}{\sqrt{\pi}} \lim_{x \to \infty} \frac{\frac{-1}{\sqrt{2q}} \exp\left(\frac{-x^2}{2q}\right)}{-x^{-2} \exp\left(\frac{-x^2}{2q}\right) - \frac{1}{q} \exp\left(\frac{-x^2}{2q}\right)} = 4 \frac{\sqrt{2q}}{\sqrt{\pi}}$$

by L'Hôspital's rule. For the second summand of the last display in (A.3) note that

$$\left|\frac{\sqrt{\pi}}{8} A_2(x)\right| \le \sum_{k=2}^{\infty} kx \int_{\frac{kx}{\sqrt{2q}}}^{\infty} \exp(-\tau^2) d\tau \cdot \exp\left(\frac{x^2}{2q}\right)$$

$$\leq \sqrt{2q} \sum_{k=2}^{\infty} \int_{\frac{kx}{\sqrt{2q}}}^{\infty} \tau \exp(-\tau^2) d\tau \cdot \exp\left(\frac{x^2}{2q}\right)$$

$$= \sqrt{\frac{q}{2}} \exp\left(\frac{-x^2}{2q}\right) \sum_{k=2}^{\infty} \exp\left(-\frac{(k^2-2)x^2}{2q}\right)$$

$$\overset{(x\geq 1)}{\leq} \sqrt{\frac{q}{2}} \exp\left(\frac{-x^2}{2q}\right) \sum_{k=2}^{\infty} \exp\left(-\frac{k^2-2}{2q}\right) = o(1) \text{ as } x \to \infty .$$

Combining the last statements with the decomposition in (A.3) yields

$$\lim_{x\to\infty} x\left(1 - F_{\mathbb{M}}(x)\right) \exp\left(\frac{x^2}{2q}\right) = 4\frac{\sqrt{2q}}{\sqrt{\pi}} . \tag{A.4}$$

For the denominator of (A.2) note that

$$\lim_{x\to\infty} F'_{\mathbb{M}}(x) \exp\left(\frac{x^2}{2q}\right) = \frac{4\sqrt{2}}{\sqrt{\pi q}} + \lim_{x\to\infty} \frac{4\sqrt{2}}{\sqrt{\pi q}} \sum_{k=2}^{\infty} (-1)^{k+1} k^2 \exp\left(-\frac{(k^2-1)x^2}{2q}\right)$$

$$= \frac{4\sqrt{2}}{\sqrt{\pi q}} , \tag{A.5}$$

where we used that for $x \geq 1$

$$\sum_{k=2}^{\infty} k^2 \exp\left(-\frac{(k^2-1)x^2}{2q}\right) \leq \exp\left(\frac{-x^2}{2q}\right) \sum_{k=2}^{\infty} k^2 \exp\left(-\frac{(k^2-2)}{2q}\right) = o(1),$$

as $x \to \infty$. Using similar arguments we obtain

$$\lim_{x\to\infty} x^{-1} F''_{\mathbb{M}}(x) \exp\left(\frac{x^2}{2q}\right)$$

$$= \frac{-4\sqrt{2}}{\sqrt{\pi q}q} + \lim_{x\to\infty} \frac{4\sqrt{2}}{\sqrt{\pi q}q} \sum_{k=2}^{\infty} (-1)^k k^4 \exp\left(-\frac{x^2(k^2-1)}{2q}\right) = \frac{-4\sqrt{2}}{\sqrt{\pi q}q} . \tag{A.6}$$

Combining (A.4), (A.5) and (A.6), it follows that

$$\lim_{x\to\infty} \frac{(1 - F_{\mathbb{M}}(x))F''_{\mathbb{M}}(x)}{(F'_{\mathbb{M}}(x))^2}$$

$$= \frac{\displaystyle\lim_{x\to\infty} x(1 - F_{\mathbb{M}}(x)) \exp\left(\frac{x^2}{2q}\right) \lim_{x\to\infty} \frac{F''_{\mathbb{M}}(x)}{x} \exp\left(\frac{x^2}{2q}\right)}{\displaystyle\lim_{x\to\infty} \left(F'_{\mathbb{M}}(x) \exp\left(\frac{x^2}{2q}\right)\right)^2} = -1,$$

which completes the proof of (A.2). By definition of the maximum domain of attraction, (A.2) is equivalent to the existence of sequences $a_d > 0, b_d$ such that

$$a_d\left(\max_{h=1}^{d} M'_h - b_d\right) \overset{\mathcal{D}}{\Longrightarrow} G \text{ as } d \to \infty .$$

Our next goal is to find an explicit formula for these sequences, for which we will employ Proposition 3.3.28 in Embrechts et al. (1997). Therefore consider the distribution function

$$H(x) := 1 - 4\frac{\sqrt{2q}}{\sqrt{\pi}}x^{-1}\exp\left(-x^2/(2q)\right), \qquad x > x_0 .$$

Note that we can rewrite equation (A.4) as

$$\lim_{x\to\infty}(1 - F_{\mathbb{M}}(x))\left(4\frac{\sqrt{2q}}{\sqrt{\pi}}x^{-1}\exp\left(-x^2/(2q)\right)\right)^{-1} = 1$$

and thereby obtain the tail-equivalence

$$\lim_{x\to\infty}\frac{1 - F_{\mathbb{M}}(x)}{1 - H(x)} = 1 .$$

Thus, by Proposition 3.3.28 in Embrechts et al. (1997) we have to find sequences a_d, b_d such that $\lim_{d\to\infty} H^d(a_d^{-1}x + b_d) = F_G(x)$ for all $x \in \mathbb{R}$. Since H has Weibull-like tails, a possible choice of the norming sequences is given in Table 3.4.4 in Embrechts et al. (1997):

$$b_d = \sqrt{2q\log d} + \frac{1}{2\sqrt{2q\log d}}\left(-q\log\left(2q\log d\right) + 2q\log\sqrt{\frac{32q}{\pi}}\right)$$
$$a_d^{-1} = q(2q\log d)^{-1/2} .$$

After simplification we get the desired result

$$a_d = \sqrt{\frac{2\log d}{q}} \quad \text{and} \quad b_d = \sqrt{2q\log d} - \frac{\sqrt{q}\left(\log\log d - \log\frac{16}{\pi}\right)}{2\sqrt{2\log d}} ,$$

which finishes the proof of Lemma A.1. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

A.2. Proof of Theorem 3.5

Let $q \in (0, 1]$ and recall the definition of M_h in Theorem 3.5. By Theorem 1 from Arratia et al. (1989) in the form as presented in Lemma A.4 in Jiang (2004) we obtain for any $x \in \mathbb{R}$ the inequality

$$\left|\mathbb{P}\left(\max_{h=1}^{d} M_h \le u_d(x)\right) - \exp(-\lambda)\right| \le \left(1 \wedge \lambda^{-1}\right)(\Lambda_1 + \Lambda_2 + \Lambda_3) , \qquad (A.7)$$

with $\lambda = \sum_{h=1}^{d} \mathbb{P}(M_h > u_d(x))$ and

$$\Lambda_1 = \sum_{i=1}^{d} \sum_{\substack{1 \leq j \leq d \\ |i-j| \leq L_d}} \mathbb{P}(M_i > u_d(x)) \mathbb{P}(M_j > u_d(x)) \ ,$$

$$\Lambda_2 = \sum_{i=1}^{d} \sum_{\substack{1 \leq j \leq d \\ j \neq i, |i-j| \leq L_d}} \mathbb{P}(M_i > u_d(x), M_j > u_d(x)) \ ,$$

$$\Lambda_3 = \sum_{i=1}^{d} \mathbb{E} \left| \mathbb{P}\Big(M_i > u_d(x) \,|\, \sigma\big(M_j : |i-j| > L_d\big)\Big) - \mathbb{P}(M_i > u_d(x)) \right| \ ,$$

where $u_d(x) = x/a_d + b_d$ and $\sigma(M_j : |i - j| > L_d)$ denotes the σ-algebra generated by the set $\{M_j : |i - j| > L_d\}$. In the remainder of the proof we fix $x \in \mathbb{R}$ and due to $a_d, b_d \to \infty$ we can assume that d is sufficiently large such that $u_d(x) > 0$. Further let $M_1', M_2', \ldots$ be i.i.d. random variables with $M_1' \overset{\mathcal{D}}{=} M_1$. With Lemma A.1 we have

$$\lim_{d \to \infty} \mathbb{P}\Big(\overset{d}{\underset{h=1}{\max}} \, M_h' \leq u_d(x) \Big) = \lim_{d \to \infty} \Big(\mathbb{P}\big(M_1 \leq u_d(x)\big) \Big)^d$$
$$= \exp\big(-\exp(-x)\big) \ . \tag{A.8}$$

As $b_d \to \infty$ and $\lim_{x \to 0} \frac{x}{\log(1-x)} = -1$, (A.8) yields

$$\lambda = \sum_{h=1}^{d} \mathbb{P}(M_h > u_d(x)) = d\mathbb{P}(M_1 > u_d(x))$$
$$= -d \log\big(1 - \mathbb{P}(M_1 > u_d(x))\big)\big(1 + o(1)\big) \tag{A.9}$$
$$= -\big(\log\big(\mathbb{P}(M_1 \leq u_d(x))^d\big)\big)\big(1 + o(1)\big) \to \exp(-x) \quad \text{as } d \to \infty \ .$$

To treat Λ_1, note that (A.9) yields for $d \to \infty$

$$1 - F_{\mathbb{M}}(u_d(x)) \sim \frac{\exp(-x)}{d} \ . \tag{A.10}$$

Since $L_d = o(d)$ by assumption, we obtain

$$\Lambda_1 = 2dL_d\big(1 - F_{\mathbb{M}}(u_d(x))\big)^2 \sim 2\frac{L_d}{d} \exp(-2x) = o(1) \ . \tag{A.11}$$

To derive the asymptotic properties of Λ_2, observe the bound

$$\Lambda_2 \leq 2dL_d \max_{\substack{1 \leq i < j \leq d \\ \rho_{i,j}=0, |i-\bar{j}| \leq L_d,}} \mathbb{P}(M_i > u_d(x), M_j > u_d(x))$$
$$+ 2dL_d \max_{\substack{1 \leq i < j \leq d \\ \rho_{i,j} \neq 0, |i-\bar{j}| \leq L_d,}} \mathbb{P}(M_i > u_d(x), M_j > u_d(x)), \tag{A.12}$$

where throughout this proof we write $\rho_{i,j}$ instead of $\rho_{i,j}^{(d)}$ for simplicity. Note, that in case of $\rho_{i,j} = 0$ it holds by (A.10), as $d \to \infty$,

$$\mathbb{P}\big(M_i > u_d(x), M_j > u_d(x)\big) = \big(\mathbb{P}\big(M_1 > u_d(x)\big)\big)^2 \sim \left(\frac{\exp(-x)}{d}\right)^2 ,$$

which implies that the first summand on the right-hand side in (A.12) vanishes. To treat the other one, we follow the idea in Lemma B.11 from Jirak (2015a) and use that a comparison of the covariance structures of the two Gaussian processes yields

$$\left\{\big(W_i(t), W_j(t)\big)\right\}_{t \geq 0} \overset{\mathcal{D}}{=} \left\{\left(W_i(t), \sqrt{1-\rho_{i,j}^2}W_j'(t) + \rho_{i,j}W_i(t)\right)\right\}_{t \geq 0} ,$$

where W_j' is a standard Wiener process that is independent of W_i. Consequently, it also holds that

$$(M_i, M_j) \overset{\mathcal{D}}{=} \left(M_i, \sup_{0 \leq t \leq q} \sup_{0 \leq s \leq t} \left|\sqrt{1-\rho_{i,j}^2}\big(W_j'(t) - W_j'(s)\big) + \rho_{i,j}\big(W_i(t) - W_i(s)\big)\right|\right)$$

and by the triangle inequality

$$\sup_{0 \leq t \leq q} \sup_{0 \leq s \leq t} \left|\sqrt{1-\rho_{i,j}^2}\big(W_j'(t) - W_j'(s)\big) + \rho_{i,j}\big(W_i(t) - W_i(s)\big)\right|$$
$$\leq \sqrt{1-\rho_{i,j}^2} \sup_{0 \leq t \leq q} \sup_{0 \leq s \leq t} \left|W_j'(t) - W_j'(s)\right| + |\rho_{i,j}| \sup_{0 \leq t \leq q} \sup_{0 \leq s \leq t} |W_i(t) - W_i(s)|$$
$$\overset{\mathcal{D}}{=} \sqrt{1-\rho_{i,j}^2}M_j' + |\rho_{i,j}|M_i ,$$

where M_j' has the same distribution as M_j but is independent of M_i. Now, we obtain for $\delta \in (0,1)$ and $\varepsilon = \delta\big(\frac{1}{|\rho_{i,j}|} - 1\big)$,

$$\mathbb{P}\big(M_i > u_d(x), M_j > u_d(x)\big) \leq \mathbb{P}\left(M_i > u_d(x), \sqrt{1-\rho_{i,j}^2}M_j' + |\rho_{i,j}|M_i > u_d(x)\right)$$
$$= \int_{u_d(x)}^\infty \mathbb{P}\left(M_j' \geq \frac{u_d(x) - y|\rho_{i,j}|}{\sqrt{1-\rho_{i,j}^2}}\right) \mathbb{P}_{M_i}(dy)$$
$$\leq \int_{u_d(x)}^{u_d(x)\left(\frac{1}{|\rho_{i,j}|} - \varepsilon\right)} \mathbb{P}\left(M_j' \geq \frac{u_d(x) - y|\rho_{i,j}|}{\sqrt{1-\rho_{i,j}^2}}\right) \mathbb{P}_{M_i}(dy)$$
$$+ \mathbb{P}\left(M_i > u_d(x)\big(\tfrac{1}{|\rho_{i,j}|} - \varepsilon\big)\right) =: P_{ij}^{(1)} + P_{ij}^{(2)} .$$

First, we bound $P_{ij}^{(2)}$. To this end, recall from (A.4) that

$$\mathbb{P}(M_1 > x) \sim K x^{-1} \exp(-x^2/(2q)), \qquad x \to \infty , \tag{A.13}$$

where $K = 4\frac{\sqrt{2q}}{\sqrt{\pi}}$. By $|\rho_{i,j}| \leq \rho_+ < 1$ we obtain

$$
\begin{aligned}
P_{ij}^{(2)} &= \mathbb{P}\Big(M_i > u_d(x)\big(\tfrac{1}{|\rho_{i,j}|}(1-\delta)+\delta\big)\Big) \leq \mathbb{P}\Big(M_i > u_d(x)\big(\tfrac{1}{\rho_+}(1-\delta)+\delta\big)\Big)\\
&\sim K\frac{1}{u_d(x)\big(\tfrac{1}{\rho_+}(1-\delta)+\delta\big)} \exp\Big(-\frac{u_d(x)^2}{2q}\big(\tfrac{1}{\rho_+}(1-\delta)+\delta\big)^2\Big)\\
&\leq \frac{1}{2}\exp\Big(-\frac{u_d(x)^2}{2q}\big(\rho_+^{-1}(1-\delta)+\delta\big)^2\Big),
\end{aligned}
$$

$$(A.14)$$

where the last inequality holds for sufficiently large d depending on ρ_+ since $u_d(x) \to \infty$. As regards $P_{ij}^{(1)}$, we have

$$
\begin{aligned}
P_{ij}^{(1)} &\leq \mathbb{P}\big(M_1 > u_d(x)\big)\,\mathbb{P}\Big(M_1 > u_d(x)\frac{\varepsilon|\rho_{i,j}|}{\sqrt{1-\rho_{i,j}^2}}\Big)\\
&= \mathbb{P}\big(M_1 > u_d(x)\big)\,\mathbb{P}\Big(M_1 > u_d(x)\delta\frac{1-|\rho_{i,j}|}{\sqrt{1-\rho_{i,j}^2}}\Big)\\
&\leq \mathbb{P}\big(M_1 > u_d(x)\big)\,\mathbb{P}\Big(M_1 > u_d(x)\delta\frac{1-\rho_+}{\sqrt{1-\rho_+^2}}\Big),
\end{aligned}
$$

where we used that the map $x \mapsto (1-x)/\sqrt{1-x^2}$ is decreasing on $[0,1)$. Employing again (A.13), we conclude for sufficiently large d

$$
\begin{aligned}
&\mathbb{P}\big(M_1 > u_d(x)\big)\,\mathbb{P}\Big(M_1 > u_d(x)\delta\frac{1-\rho_+}{\sqrt{1-\rho_+^2}}\Big)\\
&\sim K^2\frac{\sqrt{1-\rho_+^2}}{u_d(x)^2\delta(1-\rho_+)}\exp\Big(-\frac{u_d(x)^2}{2q}\Big(1+\frac{\delta^2(1-\rho_+)^2}{1-\rho_+^2}\Big)\Big)\\
&\leq \frac{1}{2}\exp\Big(-\frac{u_d(x)^2}{2q}\big(1+\delta^2(1-\rho_+)^2\big)\Big).
\end{aligned}
$$

$$(A.15)$$

Combining (A.14) and (A.15) we get

$$
\mathbb{P}\big(M_i > u_d(x), M_j > u_d(x)\big) \leq \exp\Big(-\frac{u_d(x)^2}{2q}\min(f(\delta),g(\delta))\Big),
$$

where the functions f, g are defined by $f(\delta) = \big(\rho_+^{-1}(1-\delta)+\delta\big)^2$ and $g(\delta) = 1+\delta^2(1-\rho_+)^2$. Next, we will optimize this bound in $\delta \in (0,1)$. Observe that $f(0) > g(0)$ and $f(1) < g(1)$. Since f is decreasing on $[0,1]$ while g is increasing on $[0,1]$, we deduce by continuity of f and g that there exists a unique $\delta^\star \in (0,1)$ such that $f(\delta^\star) = g(\delta^\star)$. Solving this equation, we find that

$$
\delta^\star = \frac{1-\sqrt{2\rho_+^2-\rho_+^4}}{1-\rho_+-\rho_+^2+\rho_+^3} \qquad \text{and} \qquad g(\delta^\star) = 1+\frac{\Big(1-\sqrt{2\rho_+^2-\rho_+^4}\Big)^2}{(1-\rho_+^2)^2}.
$$

Thus, it follows that

$$
2dL_d \max_{\substack{1\leq i<j\leq d\\ \rho_{i,j}\neq 0,\, |i-j|\leq L_d}} \mathbb{P}\big(M_i > u_d(x), M_j > u_d(x)\big) \leq 2dL_d \exp\Big(-\frac{u_d(x)^2}{2q}g(\delta^\star)\Big).
$$

Using the definition of $u_d(x)$ we deduce for any $\Delta < g(\delta^\star) - 1$,

$$dL_d \exp\Big(-\frac{u_d(x)^2}{2q}g(\delta^\star)\Big) \leq dL_d \exp\big(-(\Delta+1)\log d\big) = d^{-\Delta}L_d,$$

provided d is sufficiently large. By the assumption in (3.5) there exists a $\Delta \in (0, g(\delta^\star) - 1)$ such that $d^{-\Delta}L_d \to 0$ as $d \to \infty$, which together with (A.12) establishes $\Lambda_2 \to 0$.

Due to $\rho_{i,j} = 0$ for $|i - j| > L_d$ we obtain that the Gaussian processes W_i and W_j are already independent whenever $|i - j| > L_d$ [see for instance Billingsley (1999)] and therefore we have that $\Lambda_3 = 0$. The assertion now follows by combining this fact with (A.7), (A.9), (A.11) and $\Lambda_2 \to 0$.

A.3. Proof of Theorem 3.6

Throughout this proof, we will work with the sequences a_d, b_d defined in (3.6) with $q = T/(T+1) = q(T)$. Recall that the detector (2.7) is based on differences of component-wise mean estimators

$$\widehat{\mu}_{m+j+1}^{m+k}(h) - \widehat{\mu}_1^{m+j}(h)$$

and we may without loss of generality assume $\mathbb{E}[X_{t,h}] = 0$ throughout the proof. First, we introduce some necessary notations. Analogously to Theorem 3.5 let $\big(W_1, \ldots, W_d\big)^\top$ denote a d-dimensional Brownian motion on the interval $[0, q(T)]$ with correlations:

$$\text{Corr}\big(W_h(t), W_i(t)\big) = \tilde{\rho}_{h,i} := \rho_{h,i} \cdot I\{|h - i| \leq L_d\} ,\qquad (A.16)$$

where $\rho_{h,i}$ denotes the long-run correlation defined in (3.4) and $L_d := d^\Delta \to \infty$ is a sequence, where

$$\Delta < \left(\frac{1 - \rho_+\sqrt{2 - \rho_+^2}}{1 - \rho_+^2}\right)^2$$

and ρ_+ is the constant from Assumption 3.4. Note that $\tilde{\rho}_{h,i}$ depends on d only through the indicator $I\{|h - i| \leq L_d\}$. Denote again by

$$M_h := \max_{t \in [0, q(T)]} W_h(t) - \min_{t \in [0, q(T)]} W_h(t) = \max_{t \in [0, q(T)]} \max_{s \in [0, t]} \big|W_h(s) - W_h(t)\big|$$

the range of W_h. For $0 < c < T$ define additionally the truncated version

$$M_h(c) = \max_{t \in [q(c), q(T)]} \max_{s \in [0, \, q(q^{-1}(t) - c)]} \big|W_h(s) - W_h(t)\big|,\qquad (A.17)$$

where $q(x) = x/(x+1)$, $q^{-1}(x) = x/(1-x)$ and consider the overall maxima of these quantities,

$$\mathcal{W}_d = \max_{h=1}^d M_h \quad \text{and} \quad \mathcal{W}_d(c) = \max_{h=1}^d M_h(c) .\qquad (A.18)$$

Recall the definition of the Gaussian statistic $\widehat{\mathcal{T}}_{m,d}^{(Z)}$ in (3.14) based on the random variables $\{\widehat{Z}_{t,h}\}_{t=1,\ldots,m+mT}^{h=1,\ldots,d}$ defined in (3.12). We introduce two additional sets of independent, centered Gaussian random vectors

$$\left\{\boldsymbol{Z}_t = \left(Z_{t,1}, \ldots, Z_{t,d}\right)^{\top}\right\}_{t=1,\ldots,m+mT} \quad \text{and} \quad \left\{\tilde{\boldsymbol{Z}}_t = \left(Z_{t,1}, \ldots, Z_{t,d}\right)^{\top}\right\}_{t=1,\ldots,m+mT}$$

with spatial covariance structures

$$\mathrm{Cov}\left(Z_{t,h}, Z_{t,i}\right) = \rho_{h,i} \quad \text{and} \quad \mathrm{Cov}\left(\widetilde{Z}_{t,h}, \widetilde{Z}_{t,i}\right) = \tilde{\rho}_{h,i} \,, \tag{A.19}$$

where $\rho_{h,i}$ and $\tilde{\rho}_{h,i}$ are the long-run correlations and truncated long-run correlations defined in (3.4) and (A.16), respectively.

Next, we define analogues of the statistic $\widehat{\mathcal{T}}_{m,d}^{(Z)}$ in (3.14) by

$$
\begin{aligned}
\mathcal{T}_{m,d}^{(Z)} &:= \max_{h=1}^{d} \max_{k=1}^{Tm} \max_{j=0}^{k-1} \frac{(k-j)w(k/m)}{\sqrt{m}} \left| z_{m+j+1}^{m+k}(h) - z_1^{m+j}(h) \right| \,, \\
\widetilde{\mathcal{T}}_{m,d}^{(Z)} &:= \max_{h=1}^{d} \max_{k=1}^{Tm} \max_{j=0}^{k-1} \frac{(k-j)w(k/m)}{\sqrt{m}} \left| \widetilde{z}_{m+j+1}^{m+k}(h) - \widetilde{z}_1^{m+j}(h) \right| \,,
\end{aligned}
\tag{A.20}
$$

where

$$z_i^j(h) := \frac{1}{j-i+1} \sum_{t=i}^{j} Z_{t,h} \quad \text{and} \quad \widetilde{z}_i^j(h) := \frac{1}{j-i+1} \sum_{t=i}^{j} \widetilde{Z}_{t,h} \,. \tag{A.21}$$

For a constant $0 < c < T$, such that $cm \in \mathbb{N}$, we will now consider truncated versions of the statistics $\mathcal{T}_{m,d}, \mathcal{T}_{m,d}^{(Z)}, \widehat{\mathcal{T}}_{m,d}^{(Z)}$ and $\widetilde{\mathcal{T}}_{m,d}^{(Z)}$ defined by

$$
\begin{aligned}
\mathcal{T}_{m,d}(c) &:= \max_{h=1}^{d} \max_{k=cm+1}^{Tm} \max_{j=0}^{k-cm-1} \frac{(k-j)w(k/m)}{\sqrt{m}\sigma_h} \left| \widehat{\mu}_{m+j+1}^{m+k}(h) - \widehat{\mu}_1^{m+j}(h) \right| \,, \\
\mathcal{T}_{m,d}^{(Z)}(c) &:= \max_{h=1}^{d} \max_{k=cm+1}^{Tm} \max_{j=0}^{k-cm-1} \frac{(k-j)w(k/m)}{\sqrt{m}} \left| z_{m+j+1}^{m+k}(h) - z_1^{m+j}(h) \right| \,, \\
\widehat{\mathcal{T}}_{m,d}^{(Z)}(c) &:= \max_{h=1}^{d} \max_{k=cm+1}^{Tm} \max_{j=0}^{k-cm-1} \frac{(k-j)w(k/m)}{\sqrt{m}} \left| \widehat{z}_{m+j+1}^{m+k}(h) - \widehat{z}_1^{m+j}(h) \right| \,, \\
\widetilde{\mathcal{T}}_{m,d}^{(Z)}(c) &:= \max_{h=1}^{d} \max_{k=cm+1}^{Tm} \max_{j=0}^{k-cm-1} \frac{(k-j)w(k/m)}{\sqrt{m}} \left| \widetilde{z}_{m+j+1}^{m+k}(h) - \widetilde{z}_1^{m+j}(h) \right| \,.
\end{aligned}
\tag{A.22}
$$

Finally, recall the definition

$$u_d(x) = x/a_d + b_d \,, \qquad x \in \mathbb{R} \,, \tag{A.23}$$

with the sequences a_d and b_d given by (3.6) with the adaptation $q = q(T)$.

The proof of Theorem 3.6 is now split into the following five Lemmas. If these are proven, then the claim is a consequence of Theorem 3.5.

Lemma A.2 (Truncation) *For any sufficiently small constant $t_0 > 0$ and for fixed $x \in \mathbb{R}$ we have*

$$\left| \mathbb{P}\Big(\mathcal{T}_{m,d} \leq u_d(x) \Big) - \mathbb{P}\Big(\mathcal{T}_{m,d}(t_0) \leq u_d(x) \Big) \right| = o(1) \quad as\ m, d \to \infty \ .$$

Lemma A.3 (Gaussian approximation) *For $t_0 > 0$ it holds that*

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\big(\mathcal{T}_{m,d}(t_0) \leq x \big) - \mathbb{P}\big(\mathcal{T}_{m,d}^{(Z)}(t_0) \leq x \big) \right| = o(1) \quad as\ m, d \to \infty \ .$$

Lemma A.4 (Relaxation of correlation structure) *For $t_0 > 0$ it holds that*

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\big(\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) \leq x \big) - \mathbb{P}\big(\mathcal{T}_{m,d}^{(Z)}(t_0) \leq x \big) \right| = o(1) \quad as\ m, d \to \infty \ .$$

Lemma A.5 (Discretization of limit process) *For $t_0 > 0$ and fixed $x \in \mathbb{R}$ it holds that*

$$\left| \mathbb{P}\big(\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) \leq u_d(x) \big) - \mathbb{P}\big(\mathcal{W}_d(t_0) \leq u_d(x) \big) \right| = o(1) \quad as\ m, d \to \infty \ .$$

Lemma A.6 (Removing truncation) *For fixed $x \in \mathbb{R}$ and any sufficiently small constant $t_0 > 0$ it holds that*

$$\left| \mathbb{P}\big(\mathcal{W}_d(t_0) \leq u_d(x) \big) - \mathbb{P}\big(\mathcal{W}_d \leq u_d(x) \big) \right| = o(1) \quad as\ m, d \to \infty \ .$$

Proof of Lemma A.2. First, note that

$$\mathcal{T}_{m,d} = \max_{h=1}^{d} \max_{k=1}^{Tm} \max_{j=0}^{k-1} \frac{(k-j)w(k/m)}{\sqrt{m}\sigma_h} \left| \widehat{\mu}_{m+j+1}^{m+k}(h) - \widehat{\mu}_1^{m+j}(h) \right|$$

$$= \max \left\{ \mathcal{T}_{m,d}(t_0) \ , \max_{h=1}^{d} \max_{k=t_0 m+1}^{Tm} \max_{j=k-t_0 m}^{k-1} \frac{(k-j)w(k/m)}{\sqrt{m}\sigma_h} \left| \widehat{\mu}_{m+j+1}^{m+k}(h) - \widehat{\mu}_1^{m+j}(h) \right|, \right.$$

$$\left. \max_{h=1}^{d} \max_{k=1}^{t_0 m} \max_{j=0}^{k-1} \frac{(k-j)w(k/m)}{\sqrt{m}\sigma_h} \left| \widehat{\mu}_{m+j+1}^{m+k}(h) - \widehat{\mu}_1^{m+j}(h) \right| \right\} \ .$$

Hence, we obtain

$$\left| \mathbb{P}\Big(\mathcal{T}_{m,d}(t_0) \leq u_d(x) \Big) - \mathbb{P}\Big(\mathcal{T}_{m,d} \leq u_d(x) \Big) \right| \leq P_1(x) + P_2(x) \ , \qquad \text{(A.24)}$$

where

$$P_1(x) = \mathbb{P}\Big(\max_{h=1}^{d} \max_{k=t_0 m+1}^{Tm} \max_{j=k-t_0 m}^{k-1} \frac{(k-j)}{\sqrt{m}\sigma_h} \left| \widehat{\mu}_{m+j+1}^{m+k}(h) - \widehat{\mu}_1^{m+j}(h) \right| \geq u_d(x) \Big) \ ,$$

$$P_2(x) = \mathbb{P}\Big(\max_{h=1}^{d} \max_{k=1}^{t_0 m} \max_{j=0}^{k-1} \frac{(k-j)}{\sqrt{m}\sigma_h} \left| \widehat{\mu}_{m+j+1}^{m+k}(h) - \widehat{\mu}_1^{m+j}(h) \right| \geq u_d(x) \Big) \ .$$

and we additionally used that $w(k/m) \leq 1$. We will treat the summands on the right-hand side of the last display separately. For the term $P_1(x)$ note that

$$
\mathbb{P}\left(\max_{k=t_0 m+1}^{Tm} \max_{j=k-t_0 m}^{k-1} \frac{(k-j)}{\sqrt{m}\sigma_h} \left| \widehat{\mu}_{m+j+1}^{m+k}(h) - \widehat{\mu}_1^{m+j}(h) \right| \geq u_d(x) \right)
$$
$$
\leq \mathbb{P}\left(\max_{k=t_0 m+1}^{Tm} \max_{j=k-t_0 m}^{k-1} \frac{(k-j)}{\sqrt{m}\sigma_h} \left| \widehat{\mu}_{m+j+1}^{m+k}(h) \right| \geq \frac{u_d(x)}{2} \right) \qquad \text{(A.25)}
$$
$$
+ \mathbb{P}\left(\max_{k=t_0 m+1}^{Tm} \max_{j=k-t_0 m}^{k-1} \frac{(k-j)}{\sqrt{m}\sigma_h} \left| \widehat{\mu}_1^{m+j}(h) \right| \geq \frac{u_d(x)}{2} \right) .
$$

Using stationarity and Assumption 3.3 (TD2), we have

$$
\mathbb{P}\left(\max_{k=t_0 m+1}^{Tm} \max_{j=k-t_0 m}^{k-1} \frac{(k-j)}{\sqrt{m}\sigma_h} \left| \widehat{\mu}_{m+j+1}^{m+k}(h) \right| \geq \frac{u_d(x)}{2} \right)
$$
$$
\leq \sum_{k=t_0 m+1}^{Tm} \mathbb{P}\left(\max_{j=k-t_0 m}^{k-1} \left| \sum_{i=m+j+1}^{m+k} X_{i,h} \right| \geq \frac{\sqrt{m}c_\sigma u_d(x)}{2} \right) .
$$

Observing (3.6) and Lemma B.1, we obtain the following bound for the last display, which holds uniformly for $1 \leq h \leq d$

$$
C_p \frac{Tt_0 m^{2-p/2}}{c_\sigma^p u_d(x)^p} + C_p Tm \exp\left(-c_p \frac{c_\sigma^2 u_d(x)^2}{4t_0} \right) \lesssim \frac{m^{2-p/2}}{(\log(d))^{p/2}} + md^{-\tilde{C}_p/t_0} ,
$$

where $\tilde{C}_p > 0$ is a sufficiently small constant. The second summand on the right-hand side of (A.25) can be estimated similarly, that is

$$
\mathbb{P}\left(\max_{k=t_0 m+1}^{Tm} \max_{j=k-t_0 m}^{k-1} \frac{(k-j)}{\sqrt{m}\sigma_h} \left| \widehat{\mu}_1^{m+j}(h) \right| \geq \frac{u_d(x)}{2} \right)
$$
$$
\leq \mathbb{P}\left(\max_{j=1}^{Tm+m} \left| \sum_{i=1}^{j} X_{i,h} \right| \geq \frac{c_\sigma \sqrt{m}}{t_0} \frac{u_d(x)}{2} \right)
$$
$$
\lesssim C_p \frac{t_0^p (T+1) m^{1-p/2}}{c_\sigma^p u_d(x)^p} + C_p \exp\left(-c_p \frac{c_\sigma^2 u_d(x)^2}{4t_0^2(T+1)} \right) \lesssim \frac{m^{1-p/2}}{(\log(d))^{p/2}} + d^{-\tilde{C}_p/t_0^2} ,
$$

where $\tilde{C}_p > 0$ is again a sufficiently small constant. Hence, we obtain by Assumption 3.1, (A.25) (observing $p > 2D + 4$) that

$$
P_1(x) \lesssim \frac{dm^{2-p/2}}{(\log(d))^{p/2}} + md^{1-\frac{\tilde{C}_p}{t_0}}
$$
$$
\lesssim \frac{m^{D+2-p/2}}{(\log(d))^{p/2}} + m^{1+\left(1-\tilde{C}_p/t_0\right)/C_D} = o(1) \qquad \text{(A.26)}
$$

if $t_0 > 0$ is chosen sufficiently small. Analogously, we obtain for the second summand on the right-hand side of (A.24) with a possibly smaller constant $t_0 > 0$, that

$$
P_2(x) = o(1) , \qquad \text{(A.27)}
$$

where we have used the following two inequalities which are a consequence of Lemma B.1

$$\mathbb{P}\left(\max_{k=1}^{t_0 m} \max_{j=0}^{k-1} \frac{(k-j)}{\sqrt{m}\sigma_h} \left| \hat{\mu}_{m+j+1}^{m+k}(h) \right| \geq \frac{u_d(x)}{2} \right)$$

$$\leq \sum_{k=1}^{t_0 m} \mathbb{P}\left(\max_{j=0}^{k-1} \left| \sum_{i=m+j+1}^{m+k} X_{i,h} \right| \geq \sqrt{m}c_\sigma \frac{u_d(x)}{2} \right)$$

$$\leq \sum_{k=1}^{t_0 m} \mathbb{P}\left(\max_{j=0}^{t_0 m-1} \left| \sum_{i=m+j+1}^{m+t_0 m} X_{i,h} \right| \geq \sqrt{m}c_\sigma \frac{u_d(x)}{2} \right)$$

$$\lesssim C_p \frac{t_0^2 m^{2-p/2}}{c_\sigma^p (\log(d))^{p/2}} + C_p t_0 m \exp\left(-c_p \frac{c_\sigma^2 u_d(x)^2}{4 t_0} \right) \lesssim \frac{m^{2-p/2}}{(\log(d))^{p/2}} + m d^{-\tilde{C}_p/t_0}$$

and

$$\mathbb{P}\left(\max_{k=1}^{t_0 m} \max_{j=0}^{k-1} \frac{(k-j)}{\sqrt{m}\sigma_h} \left| \hat{\mu}_1^{m+j}(h) \right| \geq \frac{u_d(x)}{2} \right)$$

$$\leq \mathbb{P}\left(\max_{j=1}^{t_0 m+m} \left| \sum_{i=1}^{j} X_{i,h} \right| \geq \frac{\sqrt{m}c_\sigma}{t_0} \frac{u_d(x)}{4} \right)$$

$$\lesssim C_p \frac{(t_0+1)t_0^p m^{1-p/2}}{c_\sigma^p (\log(d))^{p/2}} + C_p \exp\left(-c_p \frac{c_\sigma^2 u_d(x)^2}{4(t_0+1)t_0^2(T+1)} \right)$$

$$\lesssim \frac{m^{1-p/2}}{(\log(d))^{p/2}} + d^{-\frac{\tilde{C}_p}{(t_0+1)t_0^2}} \ .$$

Combining (A.26) and (A.27) the assertion of Lemma A.2 now follows from (A.24).

Proof of Lemma A.3. We will use a Gaussian Approximation provided in Corollary 2.2 of Zhang and Cheng (2018). For this purpose we introduce the notation

$$v_{m,k,j,h} := \frac{(k-j)w(k/m)}{\sigma_h \sqrt{m}} \left(\hat{\mu}_{m+j+1}^{m+k}(h) - \hat{\mu}_1^{m+j}(h) \right) ,$$

with $k = t_0 m + 1, \ldots, Tm$; $j = 0 \ldots, k - t_0 m - 1$ and $h = 1, \ldots, d$. We stack all these quantities together in one vector

$$\boldsymbol{V}_+ := (v_{m,t_0 m+1,0,1}, v_{m,t_0 m+2,0,1}, v_{m,t_0 m+2,1,1}, \ldots,$$

$$v_{m,Tm,Tm-t_0 m-1,1}, v_{m,t_0 m+1,0,2}, \ldots, v_{m,Tm,Tm-t_0 m-1,d})^\top .$$

Next define the vector

$$\boldsymbol{V} = (V_1, V_2, \ldots, V_{d_{\boldsymbol{V}}})^\top := \left(\boldsymbol{V}_+^\top, -\boldsymbol{V}_+^\top \right)^\top$$

and denote its dimension by $d_{\boldsymbol{V}}$. Observe that by construction the identity

$$\max_{i=1}^{d_{\boldsymbol{V}}} V_i = \mathcal{T}_{m,d}(t_0) = \max_{h=1}^{d} \max_{k=mt_0+1}^{Tm} \max_{j=0}^{k-mt_0-1} \frac{(k-j)w(k/m)}{\sigma_h \sqrt{m}} \left| \hat{\mu}_{m+j+1}^{m+k}(h) - \hat{\mu}_1^{m+j}(h) \right|$$

holds, where we use the fact that $\boldsymbol{V}$ contains both, the positive and negative version of all random variables which appear in the maximum in the definition of the statistic $\mathcal{T}_{m,d}(t_0)$. Further note that the dimension of $\boldsymbol{V}$ is bounded by

$$d_{\boldsymbol{V}} \le 2d(Tm)^2. \tag{A.28}$$

By the construction above each component V_i corresponds either to $v_{m,k,j,h}$ or to $-v_{m,k,j,h}$ for some combination k, j, h. Hence, it can be represented by

$$V_i = \frac{1}{\sqrt{m}} \sum_{t=1}^{m(T+1)} X_{t,i}^*$$

with

$$X_{t,i}^* = \begin{cases} \dfrac{a_{t,m,k,j}}{\sigma_h} X_{t,h} & \text{for } 1 \le i \le d_{\boldsymbol{V}}/2 , \\[3mm] -\dfrac{a_{t,m,k,j}}{\sigma_h} X_{t,h} & \text{for } d_{\boldsymbol{V}}/2 + 1 \le i \le d_{\boldsymbol{V}} , \end{cases} \tag{A.29}$$

where the indices k, j, h correspond to i according to the construction of the vector $\boldsymbol{V}$ and the coefficients $a_{t,m,k,j}$ are given by

$$a_{t,m,k,j} = \begin{cases} 0 & \text{if } t > m + k , \\[2mm] a_{m,k}^{(1)} := w(k/m) & \text{if } m + j < t \le m + k , \\[3mm] a_{m,k,j}^{(2)} := -\dfrac{(k-j)w(k/m)}{(m+j)} & \text{if } t \le m + j . \end{cases} \tag{A.30}$$

Using the fact $w(k/m) = 1/(1 + k/m)$ and $1 \le k \le mT$, we obtain

$$\frac{1}{T+1} \le w(k/m) = a_{m,k}^{(1)} \le 1 \tag{A.31}$$

and as $t_0 m \le k - j$ and $j \le k \le Tm$ it follows that

$$\frac{t_0}{(T+1)^2} \le |a_{m,k,j}^{(2)}| = \left| \frac{(k-j)w(k/m)}{m+j} \right| \le T , \tag{A.32}$$

which yields by definition of $a_{t,m,k,j}$ in (A.30) the upper bound

$$|a_{t,m,k,j}| \le T_+ := \max\{T, 1\} . \tag{A.33}$$

Moreover, the temporal dependence structure of the $d_{\boldsymbol{V}}$-dimensional time series

$$\left\{ (X_{t,1}^*, \ldots, X_{1,d_{\boldsymbol{V}}}^*)^\top \right\}_{t \in \mathbb{Z}}$$

still satisfies the concept of physical dependence as

$$X_{t,i}^* = g_{t,i}^*(\varepsilon_t, \varepsilon_{t-1}, \ldots)$$

with

$$g^*_{m,t,i}(\varepsilon_t, \varepsilon_{t-1}, \dots) := \begin{cases} \dfrac{a_{t,m,k,j}}{\sigma_h} g_h(\varepsilon_t, \varepsilon_{t-1}, \dots) & \text{for } 1 \le i \le d_{\boldsymbol{V}}/2 , \\[3mm] -\dfrac{a_{t,m,k,j}}{\sigma_h} g_h(\varepsilon_t, \varepsilon_{t-1}, \dots) & \text{for } d_{\boldsymbol{V}}/2 + 1 \le i \le d_{\boldsymbol{V}} , \end{cases}$$

$$(A.34)$$

where the indices k, j, h correspond to i according to the construction of the vector $\boldsymbol{V}$.

In the following let $(\dot{V}_1^{(z)}, \dots, \dot{V}_{d_{\boldsymbol{V}}}^{(z)})^\top$ denote a centered Gaussian distributed vector with the same covariance structure as $\boldsymbol{V}$. Next, recall the definition of the Gaussian random variables $\{Z_{t,h}\}_{t=1,\dots,m+mT}^{h=1,\dots,d}$ in (A.19) and let

$$Z^*_{t,i} = \begin{cases} a_{t,m,k,j} Z_{t,h} & \text{for } 1 \le i \le d_{\boldsymbol{V}}/2 , \\[3mm] -a_{t,m,k,j} Z_{t,h} & \text{for } d_{\boldsymbol{V}}/2 + 1 \le i \le d_{\boldsymbol{V}} . \end{cases}$$

Further define the vector $\boldsymbol{V}^{(z)} = (V_1^{(z)}, \dots, V_{d_{\boldsymbol{V}}}^{(z)})^\top$ by

$$V_i^{(z)} := \frac{1}{\sqrt{m}} \sum_{t=1}^{m(T+1)} Z^*_{t,i} \quad i = 1, \dots, d_{\boldsymbol{V}} .$$

We now proceed as follows:

Step 1: Show that for some (sufficiently small) constant $\tilde{C}_1 > 0$

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(\max_{i=1}^{d_{\boldsymbol{V}}} \dot{V}_i^{(z)} \le x \right) - \mathbb{P}\left(\max_{i=1}^{d_{\boldsymbol{V}}} V_i^{(z)} \le x \right) \right| \lesssim m^{-\tilde{C}_1}.$$

Step 2: Establish that for some (sufficiently small) constant $\tilde{C}_2 > 0$

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(\max_{i=1}^{d_{\boldsymbol{V}}} V_i \le x \right) - \mathbb{P}\left(\max_{i=1}^{d_{\boldsymbol{V}}} \dot{V}_i^{(z)} \le x \right) \right| \lesssim m^{-\tilde{C}_2} . \qquad (A.35)$$

If both steps have been proven, the claim of Lemma A.3 follows from the identity

$$\max_{i=1}^{d_{\boldsymbol{V}}} V_i^{(z)} = \mathcal{T}_{m,d}^{(Z)}(t_0) .$$

Proof of Step 1: As we aim to compare the maxima of the two Gaussian distributed vectors $\boldsymbol{V}^{(z)}$ and $\dot{\boldsymbol{V}}^{(z)}$ we will apply Lemma B.3. Therefore, we analyze the covariance structures of $\boldsymbol{V}^{(z)}$ and $\dot{\boldsymbol{V}}^{(z)}$ [or equivalently $\boldsymbol{V}$]. Let $i_1, i_2 \in \{1, \dots, d_{\boldsymbol{V}}/2\}$ with corresponding indices h_1, j_1, k_1 and h_2, j_2, k_2 according to equation (A.29). For the calculation we assume without loss of generality that $j_1 \le j_2$ and use the notation $k_{\min} = \min\{k_1, k_2\}$, $k_{\max} = \max\{k_1, k_2\}$

and $j_2 \wedge k_1 = \min\{j_2, k_1\}$. Further we use the convention $\sum_{i=j}^{k} \beta_i = 0$, whenever $k < j$. For the covariance of the components of the vector $\boldsymbol{V}^{(z)}$ note that temporal independence yields

$$
\begin{aligned}
\mathrm{Cov}\left(V_{i_1}^{(z)}, V_{i_2}^{(z)}\right) &= \frac{1}{m} \sum_{t=1}^{m+k_{\min}} \mathrm{Cov}\left(a_{t,m,k_1,j_1} Z_{t,h_1},\, a_{t,m,k_2,j_2} Z_{t,h_2}\right) \\
&= \frac{1}{m} \sum_{t=1}^{m+j_1} \mathrm{Cov}\left(a_{t,m,k_1,j_1} Z_{t,h_1},\, a_{t,m,k_2,j_2} Z_{t,h_2}\right) \\
&\quad + \frac{1}{m} \sum_{t=m+j_1+1}^{m+(j_2 \wedge k_1)} \mathrm{Cov}\left(a_{t,m,k_1,j_1} Z_{t,h_1},\, a_{t,m,k_2,j_2} Z_{t,h_2}\right) \\
&\quad + \frac{1}{m} \sum_{t=m+(j_2 \wedge k_1)+1}^{m+k_{\min}} \mathrm{Cov}\left(a_{t,m,k_1,j_1} Z_{t,h_1},\, a_{t,m,k_2,j_2} Z_{t,h_2}\right).
\end{aligned}
$$

Using the definition in (A.30) and (A.19), we obtain

$$
\begin{aligned}
\mathrm{Cov}\left(V_{i_1}^{(z)}, V_{i_2}^{(z)}\right) &= \frac{a_{m,k_1,j_1}^{(2)} a_{m,k_2,j_2}^{(2)}}{m\sigma_{h_1}\sigma_{h_2}} (m+j_1)\gamma_{h_1,h_2} \\
&\quad + \frac{a_{m,k_1}^{(1)} a_{m,k_2,j_2}^{(2)}}{m\sigma_{h_1}\sigma_{h_2}} \left((j_2 \wedge k_1) - j_1\right)\gamma_{h_1,h_2} \\
&\quad + \frac{a_{m,k_1}^{(1)} a_{m,k_2}^{(1)}}{m\sigma_{h_1}\sigma_{h_2}} (k_{\min} - j_2)\gamma_{h_1,h_2}\, I\{j_2 < k_{\min}\}.
\end{aligned}
\tag{A.36}
$$

Similar calculations also yield

$$
\mathrm{Var}\left(V_{i_1}^{(z)}\right) = \mathrm{Cov}\left(V_{i_1}^{(z)}, V_{i_1}^{(z)}\right) = \left(a_{m,k_1,j_1}^{(2)}\right)^2 \frac{(m+j_1)}{m} + \left(a_{m,k_1}^{(1)}\right)^2 \frac{k_1 - j_1}{m}
$$

and from (A.31) and (A.32) it follows that

$$
\frac{t_0^2}{(T+1)^4} + \frac{t_0}{(T+1)^2} \leq \mathrm{Var}\left(V_{i_1}^{(z)}\right) \leq T^3 + T.
\tag{A.37}
$$

By the same arguments we obtain for the covariance structure of the components of the vector $\dot{\boldsymbol{V}}^{(z)}$ [note that we cannot use temporal independence here]:

$$
\begin{aligned}
\mathrm{Cov}\left(\dot{V}_{i_1}^{(z)}, \dot{V}_{i_2}^{(z)}\right) &= \mathrm{Cov}\left(\frac{1}{\sqrt{m}} \sum_{t=1}^{m+k_1} X_{t,i_1}^*,\, \frac{1}{\sqrt{m}} \sum_{t=1}^{m+k_2} X_{t,i_2}^*\right) \\
&= \sum_{\ell=1}^{4} \mathrm{Cov}\left(S_\ell^{(1)}, S_\ell^{(2)}\right) + \sum_{\substack{\ell,j=1 \\ \ell \neq j}}^{4} \mathrm{Cov}\left(S_\ell^{(1)}, S_j^{(2)}\right),
\end{aligned}
\tag{A.38}
$$

where the terms in the last line are defined for $u = 1, 2$ by

$$S_1^{(u)} = \frac{1}{\sqrt{m}} \sum_{t=1}^{m+j_1} X_{t,i_u}^*, \quad S_2^{(u)} = \frac{1}{\sqrt{m}} \sum_{t=m+j_1+1}^{m+(j_2 \wedge k_1)} X_{t,i_u}^*,$$

$$S_3^{(u)} = \frac{1}{\sqrt{m}} \sum_{t=m+(j_2 \wedge k_1)+1}^{m+k_{\min}} X_{t,i_u}^*, \quad S_4^{(u)} = \frac{1}{\sqrt{m}} \sum_{t=m+k_{\min}+1}^{m+k_{\max}} X_{t,i_u}^*.$$

We will now treat the two sums on the right-hand side of (A.38) separately and show that the first sum is close to $\mathrm{Cov}\left(V_{i_1}^{(z)}, V_{i_2}^{(z)}\right)$, while the second vanishes sufficiently fast. Using that by construction, either $S_4^{(1)} = 0$ or $S_4^{(2)} = 0$, we obtain that

$$\sum_{\ell=1}^{4} \mathrm{Cov}\left(S_\ell^{(1)}, S_\ell^{(2)}\right) = \frac{a_{m,k_1,j_1}^{(2)} a_{m,k_2,j_2}^{(2)}}{m\sigma_{h_1}\sigma_{h_2}} \sum_{t=1}^{m+j_1} \sum_{s=1}^{m+j_1} \mathrm{Cov}\left(X_{t,h_1}, X_{s,h_2}\right)$$

$$+ \frac{a_{m,k_2}^{(1)} a_{m,k_2,j_2}^{(2)}}{m\sigma_{h_1}\sigma_{h_2}} \sum_{t=m+j_1+1}^{m+(j_2 \wedge k_1)} \sum_{s=m+j_1+1}^{m+(j_2 \wedge k_1)} \mathrm{Cov}\left(X_{t,h_1}, X_{s,h_2}\right)$$

$$+ \frac{a_{m,k_1}^{(1)} a_{m,k_2}^{(1)}}{m\sigma_{h_1}\sigma_{h_2}} \sum_{t=m+(j_2 \wedge k_1)+1}^{m+k_{\min}} \sum_{s=m+(j_2 \wedge k_1)+1}^{m+k_{\min}} \mathrm{Cov}\left(X_{t,h_1}, X_{s,h_2}\right)$$

$$= \frac{a_{m,k_1,j_1}^{(2)} a_{m,k_2,j_2}^{(2)}}{m\sigma_{h_1}\sigma_{h_2}} \sum_{t=-m-j_1}^{m+j_1} \left(m + j_1 - |t|\right)\phi_{t,h_1,h_2}$$

$$+ \frac{a_{m,k_2}^{(1)} a_{m,k_2,j_2}^{(2)}}{m\sigma_{h_1}\sigma_{h_2}} \sum_{t=-(j_2 \wedge k_1)+j_1}^{(j_2 \wedge k_1)-j_1} \left((j_2 \wedge k_1) - j_1 - |t|\right)\phi_{t,h_1,h_2}$$

$$+ \frac{a_{m,k_1}^{(1)} a_{m,k_2}^{(1)}}{m\sigma_{h_1}\sigma_{h_2}} I\{k_{\min} > j_2 \wedge k_1\} \sum_{t=-k_{\min}+(j_2 \wedge k_1)}^{k_{\min}-(j_2 \wedge k_1)} \left(k_{\min} - (j_2 \wedge k_1) - |t|\right)\phi_{t,h_1,h_2},$$

where we used the notation $\phi_{t,h_1,h_2} := \mathrm{Cov}(X_{0,h_1}, X_{t,h_2})$. Combining the bounds in (A.31) and (A.32) with (A.36) and Assumption 3.3 (TD2), we deduce that

$$\left| \mathrm{Cov}\left(V_{i_1}^{(z)}, V_{i_2}^{(z)}\right) - \sum_{\ell=1}^{4} \mathrm{Cov}\left(S_\ell^{(1)}, S_\ell^{(2)}\right) \right| \leq \frac{C_{T,t_0}}{c_\sigma^2 m} \left[\sum_{t\in\mathbb{Z}} \min\{|t|, m+j_1\}|\phi_{t,h_1,h_2}| \right.$$

$$\left. + \sum_{t\in\mathbb{Z}} \min\{|t|, (j_2 \wedge k_1) - j_1\}|\phi_{t,h_1,h_2}| + \sum_{t\in\mathbb{Z}} \min\{|t|, k_{\min} - (j_2 \wedge k_1)\}|\phi_{t,h_1,h_2}| \right]$$

$$\leq \frac{3C_{T,t_0}}{c_\sigma^2 m} \sum_{t\in\mathbb{Z}} |t||\phi_{t,h_1,h_2}|,$$

where the constant C_{T,t_0} depends on T and t_0 only and we used the definition of γ_{h_1,h_2} in (3.3). Using Assumption 3.3 (TD1) and Lemma E4 from Jirak (2015b)

it follows that

$$\sup_{h_1, h_2 \in \mathbb{N}} \sum_{t \in \mathbb{Z}} |t| |\phi_{t,h_1,h_2}| < \infty \ , \tag{A.39}$$

which yields

$$\left| \text{Cov} \left(V_{i_1}^{(z)}, V_{i_2}^{(z)} \right) - \sum_{\ell=1}^{4} \text{Cov} \left(S_\ell^{(1)}, S_\ell^{(2)} \right) \right| \lesssim \frac{1}{m} \ , \tag{A.40}$$

where the involved constant is independent of i_1 and i_2 [or equivalently of j_1, j_2, k_1, k_2, h_1 and h_2]. Next, we treat the second sum on the right-hand of (A.38). For that purpose, note that for arbitrary points in time $p_1 < p_2 < p_3 < p_4$, it holds that

$$\left| \text{Cov} \left(\sum_{t=p_1}^{p_2} X_{t,h_1}, \sum_{s=p_3}^{p_4} X_{s,h_2} \right) \right| \leq \sum_{t=1}^{p_2} \sum_{s=p_2+1}^{p_4} \left| \text{Cov} \left(X_{t,h_1}, X_{s,h_2} \right) \right|$$

$$= \sum_{t=1}^{p_2} \sum_{s=p_2+1}^{p_4} |\phi_{s-t,h_1,h_2}| = \sum_{t=1}^{p_2} \sum_{s=p_2-t+1}^{p_4-t} |\phi_{s,h_1,h_2}|$$

$$\leq \sum_{t=1}^{p_2} \sum_{s=p_2-t+1}^{p_4} |\phi_{s,h_1,h_2}| = \sum_{s=1}^{p_4} \sum_{t=p_2-s+1}^{p_2} |\phi_{s,h_1,h_2}| = \sum_{s=1}^{p_4} s |\phi_{s,h_1,h_2}| \ . \tag{A.41}$$

Using the upper bound for the coefficients $a_{t,m,k,j}$ in (A.33), the uniform bound in (A.39) and that all the pairs of the sums under consideration are non-overlapping as treated above in (A.41), we obtain directly that

$$\sum_{\substack{\ell,j=1 \\ i \neq j}}^{4} \text{Cov} \left(S_\ell^{(1)}, S_j^{(2)} \right) \lesssim \frac{1}{m} \ , \tag{A.42}$$

where the constant is again independent of i_1, i_2. Combining the estimates (A.40) and (A.42), we conclude

$$\Delta_m := \max_{i_1, i_2=1}^{d_{\boldsymbol{V}}} \left| \text{Cov} \left(V_{i_1}^{(z)}, V_{i_2}^{(z)} \right) - \text{Cov} \left(\dot{V}_{i_1}^{(z)}, \dot{V}_{i_2}^{(z)} \right) \right| \lesssim \frac{1}{m} \ . \tag{A.43}$$

Due to (A.37) we can now apply Lemma B.3, which gives

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left(\max_{i=1}^{d_{\boldsymbol{V}}} V_i^{(z)} \leq x \right) - \mathbb{P} \left(\max_{i=1}^{d_{\boldsymbol{V}}} \dot{V}_i^{(z)} \leq x \right) \right| \lesssim \Delta_m^{1/3} \max \left\{ 1, \log \left(d_{\boldsymbol{V}} / \Delta_m \right) \right\}^{2/3}$$

$$\lesssim \max \left\{ \Delta_m^{1/2}, \Delta_m^{1/2} \left| \log d_{\boldsymbol{V}} \right| + \Delta_m^{1/2} \left| \log \Delta_m \right| \right\}^{2/3} \ .$$

Using (A.28) and Assumption 3.1 the assertion of Step 1 follows.

Proof of Step 2: Corollary 2.2 of Zhang and Cheng (2018) yields the Gaussian approximation in (A.35) if the following three conditions hold uniformly in t and i (or equivalently in t, k, j, h).

(i) There exist a constant $b \in [0, 1/11)$ and a deterministic sequence $B_m^* \lesssim m^{(3-17b)/8}$ such that

$$d_{\boldsymbol{V}} \lesssim \exp((Tm)^b) \quad \text{and} \quad \mathbb{E}\left[\exp\left(|X_{t,i}^*|/B_m^*\right)\right] \leq C_e\,,$$

where $C_e > 1$ is the constant from Assumption 3.2 (S1).

(ii) With β as in Assumption 3.3 (TD1) it holds

$$\sum_{\ell=u}^{\infty} \sup_{t \in \mathbb{Z}} \left\| g_{t,i}^*(\varepsilon_t, \varepsilon_{t-1}, \ldots) - g_{t,i}^*(\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-\ell+1}, \varepsilon_{t-\ell}', \varepsilon_{t-\ell-1}, \ldots) \right\|_p \lesssim \beta^u,$$

where ε_{t-l}' is an independent copy of $\varepsilon_{t-\ell}$.

(iii) There exist positive constants c_1, c_2 such that $c_1 \leq \mathrm{Var}\left(V_i\right) \leq c_2$.

Therefore the proof of Lemma A.3 is completed by establishing these conditions. Proof of (i): By (A.28) and Assumption 3.1 (D1) the inequality $d_{\boldsymbol{V}} \lesssim \exp((Tm)^b)$ holds for any $b > 0$. Due to Assumption 3.2 and the upper bound on $|a_{t,m,k,j}|$ in (A.33) we obtain that

$$|X_{t,i}^*| \overset{\mathcal{D}}{=} \frac{|a_{t,m,k,j}|}{\sigma_h} |X_{1,h}| \leq \frac{T_+}{c_\sigma} |X_{1,h}|.$$

Defining $B_m^* := \dfrac{T_+}{c_\sigma} B_m \lesssim m^B$, where B_m is the sequence from Assumption 3.2 and $B < 3/8$, it follows that

$$\mathbb{E}\left[\exp\left(|X_{t,i}^*|/B_m^*\right)\right] = \mathbb{E}\left[\exp\left(|a_{t,m,k,j}||X_{t,h}|/(\sigma_h B_m^*)\right)\right]$$
$$\leq \mathbb{E}\left[\exp\left(|X_{1,h}|/B_m\right)\right] \leq C_e\,.$$

As $d_{\boldsymbol{V}} \lesssim \exp((Tm)^b)$ holds for any $b > 0$, we can choose b to be sufficiently small such that $B < (3 - 17b)/8 < 3/8$.

Proof of (ii): In view of (A.34), (A.33), Assumption 3.3 (TD1) and the stationarity of $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ we obtain

$$\sum_{\ell=u}^{\infty} \sup_{t \in \mathbb{Z}} \left\| g_{m,t,i}^*(\varepsilon_t, \varepsilon_{t-1}, \ldots) - g_{m,t,i}^*(\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-\ell+1}, \varepsilon_{t-\ell}', \varepsilon_{t-\ell-1}, \ldots) \right\|_p$$
$$= \sum_{\ell=u}^{\infty} \sup_{t \in \mathbb{Z}} \frac{|a_{t,m,k,j}|}{\sigma_h} \left\| g_h(\varepsilon_\ell, \varepsilon_{\ell-1}, \ldots) - g_h(\varepsilon_\ell, \varepsilon_{\ell-1}, \ldots, \varepsilon_1, \varepsilon_0', \varepsilon_{-1}, \ldots) \right\|_p$$
$$\leq \frac{T_+}{c_\sigma} \sum_{\ell=u}^{\infty} \vartheta_{\ell,h,p} \leq \frac{T_+ C_\vartheta}{c_\sigma} \sum_{\ell=u}^{\infty} \beta^\ell \lesssim \beta^u\,.$$

Proof of (iii): The assertion follows by combining (A.37) and (A.43). $\qquad\square$

Proof of Lemma A.4. From the proof of Lemma A.3, recall the definition of the Gaussian vector $\boldsymbol{V}^{(z)} = \left(V_1^{(z)}, \ldots, V_{d_{\boldsymbol{V}}}^{(z)}\right)^\top$, which fulfills the identity

$$\mathcal{T}_{m,d}^{(Z)}(t_0) = \max_{i=1}^{d_{\boldsymbol{V}}} V_i^{(z)}\,.$$

Applying again the vectorization technique as introduced in the proof of Lemma A.3, we can define analogously a Gaussian vector for the statistic $\widetilde{\mathcal{T}}_{m,d}^{(Z)}$ in (A.22). Recall the definition of $\tilde{z}_i^j$ in (A.21) and introduce the notation

$$\tilde{v}_{m,k,j,h} := \frac{(k-j)w(k/m)}{\sqrt{m}}\left(\tilde{z}_{m+j+1}^{m+k}(h) - \tilde{z}_1^{m+j}(h)\right), \qquad (A.44)$$

with $k = t_0 m + 1, \ldots, Tm$ and $j = 0 \ldots, k - t_0 m - 1$ and $h = 1, \ldots, d$. We stack all these quantities together in one vector, this is

$$\widetilde{\boldsymbol{V}}_+^{(z)} := (\tilde{v}_{m,t_0m+1,0,1}, \tilde{v}_{m,t_0m+2,0,1}, \tilde{v}_{m,t_0m+2,1,1}, \ldots,$$
$$\tilde{v}_{m,Tm,Tm-t_0m-1,1}, \tilde{v}_{m,t_0m+1,0,2}, \ldots, \tilde{v}_{m,Tm,Tm-t_0m-1,d})^\top.$$

Next let $\widetilde{\boldsymbol{V}}^{(z)} = \left(\left(\widetilde{\boldsymbol{V}}_+^{(z)}\right)^\top, -\left(\widetilde{\boldsymbol{V}}_+^{(z)}\right)^\top\right)^\top$ with dimension $d_{\boldsymbol{V}}$ and denote its components by

$$\widetilde{\boldsymbol{V}}^{(z)} = (\widetilde{V}_1^{(z)}, \widetilde{V}_2^{(z)}, \ldots, \widetilde{V}_{d_{\boldsymbol{V}}}^{(z)})^\top.$$

By construction of $\widetilde{\boldsymbol{V}}^{(z)}$ we have

$$\widetilde{\mathcal{T}}_{m,d}^{(Z)} = \max_{i=1}^{d_{\boldsymbol{V}}} \widetilde{V}_i^{(z)}.$$

The covariance structure of $\boldsymbol{V}^{(z)}$ was already calculated in (A.36) and is given by

$$\mathrm{Cov}\left(V_{i_1}^{(z)}, V_{i_2}^{(z)}\right) = \frac{a_{m,k_1,j_1}^{(2)} a_{m,k_2,j_2}^{(2)}}{m}(m+j_1)\rho_{h_1,h_2}$$
$$+ \frac{a_{m,k_1}^{(1)} a_{m,k_2,j_2}^{(2)}}{m}\left((j_2 \wedge k_1) - j_1\right)\rho_{h_1,h_2} \qquad (A.45)$$
$$+ \frac{a_{m,k_1}^{(1)} a_{m,k_2}^{(1)}}{m}\rho_{h_1,h_2}(k_{\min} - j_2)I\{j_2 < k_{\min}\},$$

where k_1, j_1, h_1 and k_2, j_2, h_2 are the corresponding indices to i_1 and i_2, respectively, and we use the notation $k_{\min} = \min\{k_1, k_2\}$. A similar calculation for the vector $\widetilde{\boldsymbol{V}}^{(z)}$ gives

$$\mathrm{Cov}\left(\widetilde{V}_{i_1}^{(z)}, \widetilde{V}_{i_2}^{(z)}\right) = \frac{a_{m,k_1,j_1}^{(2)} a_{m,k_2,j_2}^{(2)}}{m}(m+j_1)\tilde{\rho}_{h_1,h_2}$$
$$+ \frac{a_{m,k_1}^{(1)} a_{m,k_2,j_2}^{(2)}}{m}\left((j_2 \wedge k_1) - j_1\right)\tilde{\rho}_{h_1,h_2} \qquad (A.46)$$
$$+ \frac{a_{m,k_1}^{(1)} a_{m,k_2}^{(1)}}{m}\tilde{\rho}_{h_1,h_2}(k_{\min} - j_2)I\{j_2 < k_{\min}\}.$$

Note that by definition of the truncated correlations in (A.16) the quantities in (A.45) and (A.46) coincide, whenever $|h_1 - h_2| \le L_d$. Therefore we obtain for the maximum difference of the covariances,

$$\Delta_m := \max_{i_1,i_2=1}^{d_V} \left| \operatorname{Cov}\left(V_{i_1}^{(z)}, V_{i_2}^{(z)}\right) - \operatorname{Cov}\left(\widetilde{V}_{i_1}^{(z)}, \widetilde{V}_{i_2}^{(z)}\right) \right| \le C_T \sup_{\substack{h_1,h_2=1\ldots,d \\ |h_1-h_2|>L_d}} |\rho_{h_1,h_2}| ,$$

where C_T is a constant depending on T only, as we used that $j_1, j_2, k_1, k_2 \le mT$ and the upper bound in (A.33). Assumption 3.4 (SD1) and $L_d = d^\Delta$ now yields

$$\Delta_m \lesssim \log^{-2}(L_d) r_{L_d} = o\left(\log^{-2}(d)\right) .$$

Due to (A.37), we can apply Lemma B.3, which gives

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\max_{i=1}^{d_V} V^{(z)} \le x \right) - \mathbb{P}\left(\max_{i=1}^{d_V} \widetilde{V}_i^{(z)} \le x \right) \right| \lesssim \Delta_m^{1/3} \max\left\{1, \log\left(d_V/\Delta_m\right)\right\}^{2/3}$$

$$\le \max\left\{ \Delta_m^{1/2}, \Delta_m^{1/2} \big|\log d_V\big| + \Delta_m^{1/2} \big|\log \Delta_m\big| \right\}^{2/3}.$$

In view of (A.28) and Assumption 3.1 (D1), the proof of Lemma A.4 is completed. □

Proof of Lemma A.5. We use similar arguments as given in the proof of Lemma B.7 of Jirak (2015b). Let $\{W_h'\}_{h\in\mathbb{N}}$ denote an independent copy of the sequence of Brownian motions $\{W_h\}_{h\in\mathbb{N}}$ defined in (A.16). Recalling the notation (A.21) we obtain the representation

$$\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) = \max_{h=1}^{d} \max_{k=t_0 m+1}^{Tm} \max_{j=0}^{k-mt_0-1} \frac{1}{\sqrt{m}(1+k/m)} \left| \sum_{t=m+j+1}^{m+k} \widetilde{Z}_{t,h} - \frac{k-j}{m+j} \sum_{t=1}^{m+j} \widetilde{Z}_{t,h} \right|.$$

To investigate the quantities in the maximum we note that

$$\frac{1}{\sqrt{m}(1+k/m)} \left| \sum_{t=m+j+1}^{m+k} \widetilde{Z}_{t,h} - \frac{k-j}{m+j} \sum_{t=1}^{m+j} \widetilde{Z}_{t,h} \right|$$

$$= \frac{1}{\sqrt{m}(1+k/m)} \left| \sum_{t=1}^{m+k} \widetilde{Z}_{t,h} - \frac{m+k}{m+j} \sum_{t=1}^{m+j} \widetilde{Z}_{t,h} \right|$$

$$\stackrel{\mathcal{D}}{=} \frac{1}{1+k/m} \left| W_h(k/m+1) - \frac{m+k}{m+j} W_h(j/m+1) \right|$$

$$= \frac{1}{1+k/m} \left| W_h(\tfrac{k}{m}+1) - W_h(1) - \tfrac{m+k}{m+j}\left(W_h(\tfrac{j}{m}+1) - W_h(1) \right) - \tfrac{k-j}{m+j} W_h(1) \right|$$

$$\stackrel{\mathcal{D}}{=} \frac{1}{1+k/m} \left| W_h(k/m) - \frac{m+k}{m+j} W_h(j/m) - \frac{k-j}{m+j} W_h'(1) \right|$$

$$= \frac{1}{(1+k/m)(1+j/m)} \left| (1+j/m)\left\{ W_h(k/m) - k/m W_h'(1) \right\} \right.$$

$$- (1 + k/m)\Big\{W_h(j/m) - j/mW_h'(1)\Big\}\Big|, \tag{A.47}$$

where in all steps the correlation structure of $\{W_h\}_{h\in\mathbb{N}}$ is preserved. A calculation of the covariance kernel implies the identity (in distribution)

$$\Big\{W_h(t) - tW_h'(1)\Big\}_{t\geq 0,\, h\in\mathbb{N}} \overset{\mathcal{D}}{=} \Big\{(1+t)W_h\Big(\frac{t}{t+1}\Big)\Big\}_{t\geq 0,\, h\in\mathbb{N}}.$$

Applying this to (A.47) yields

$$\frac{1}{(1+k/m)(1+j/m)}\Big|(1+j/m)\Big\{W_h(k/m) - k/mW_h'(1)\Big\}$$
$$- (1 + \tfrac{k}{m})\Big\{W_h(\tfrac{j}{m}) - \tfrac{j}{m}W_h'(1)\Big\}\Big| \overset{\mathcal{D}}{=} \Big|W_h\Big(\frac{k}{m+k}\Big) - W_h\Big(\frac{j}{m+j}\Big)\Big|.$$

This now gives

$$\max_{k=t_0 m+1}^{Tm} \max_{j=0}^{k-mt_0-1} \frac{1}{\sqrt{m}(1+k/m)}\Big|\sum_{t=m+j+1}^{m+k} \widetilde{Z}_{t,h} - \frac{k-j}{m+j}\sum_{t=1}^{m+j} \widetilde{Z}_{t,h}\Big|$$
$$\overset{\mathcal{D}}{=} \max_{k=t_0 m+1}^{Tm} \max_{j=0}^{k-mt_0-1} \Big|W_h\Big(\frac{k}{m+k}\Big) - W_h\Big(\frac{j}{m+j}\Big)\Big|$$
$$= \max_{\substack{j,k\in\{1,\dots,Tm\}\\ k-j>mt_0}} \Big|W_h\Big(\frac{k}{m+k}\Big) - W_h\Big(\frac{j}{m+j}\Big)\Big| =: M_{h,m}(t_0),$$

which is the discrete counterpart of the random variable $M_h(t_0)$ defined in (A.17). Observing the identity

$$M_h(t_0) = \max_{t\in[q(t_0),\,q(T)]} \max_{s\in[0,\,q(q^{-1}(t)-t_0)]} \big|W_h(t) - W_h(s)\big|$$
$$= \max_{t\in[q(t_0),\,q(T)]} \max_{s\in[0,\,q^{-1}(t)-t_0]} \big|W_h(t) - W_h(q(s))\big|$$
$$= \max_{t\in[t_0,T]} \max_{s\in[0,t-t_0]} \Big|W_h\Big(\frac{t}{t+1}\Big) - W_h\Big(\frac{s}{s+1}\Big)\Big|,$$

the inequality $M_{h,m}(t_0) \leq M_h(t_0)$ already yields

$$\mathbb{P}\Big(\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) \leq u_d(x)\Big) = \mathbb{P}\Big(\max_{h=1}^{d} M_{h,m}(t_0) \leq u_d(x)\Big)$$
$$\geq \mathbb{P}\Big(\max_{h=1}^{d} M_h(t_0) \leq u_d(x)\Big) = \mathbb{P}\Big(\mathcal{W}_d(t_0) \leq u_d(x)\Big)$$

for all $x \in \mathbb{R}$. So it remains to find a suitable upper bound for

$$\mathbb{P}\Big(\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) \leq u_d(x)\Big) - \mathbb{P}\Big(\mathcal{W}_d(t_0) \leq u_d(x)\Big) \geq 0. \tag{A.48}$$

Observing the inequality (which holds for all $y \in \mathbb{R}$)

$$\mathbb{P}\Big(\mathcal{W}_d(t_0) \le u_d(x)\Big) \ge \mathbb{P}\Big(\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) \le u_d(x) - y, \big|\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) - \mathcal{W}_d(t_0)\big| < y\Big)$$

$$\ge \mathbb{P}\Big(\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) \le u_d(x) - y\Big) - \mathbb{P}\Big(\big|\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) - \mathcal{W}_d(t_0)\big| > y\Big)$$

the left-hand side in (A.48) is bounded by

$$\mathbb{P}\Big(\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) \le u_d(x)\Big) - \mathbb{P}\Big(\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) \le u_d(x) - y\Big) + \mathbb{P}\Big(\big|\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) - \mathcal{W}_d(t_0)\big| > y\Big).$$

We now choose $y_d = m^{-1/3}$. Then the claim is a consequence of the following two assertions:

(i) $\mathbb{P}\Big(\big|\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) - \mathcal{W}_d(t_0)\big| > y_d\Big) = o(1)$,

(ii) $\mathbb{P}\Big(\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) \le u_d(x)\Big) - \mathbb{P}\Big(\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) \le u_d(x) - y_d\Big) = o(1)$,

which will be proven below to complete the proof of Lemma A.5. To show (i), note that due to the time reversal and scaling properties of Brownian motions, it holds for all $k \le Tm$, $1 \le h \le d$

$$\max_{t \in [(k-1)/m, \, k/m]} \Big|W_h\Big(\frac{t}{t+1}\Big) - W_h\Big(\frac{k/m}{1 + k/m}\Big)\Big|$$

$$\stackrel{\mathcal{D}}{=} \max_{\lambda \in \big[0, \, \frac{k}{m+k} - \frac{k-1}{m+k-1}\big]} \Big|W_h\Big(\frac{k/m}{1 + k/m} - \lambda\Big) - W_h\Big(\frac{k/m}{1 + k/m}\Big)\Big|$$

$$\stackrel{\mathcal{D}}{=} \max_{\lambda \in [0, \, \frac{k}{m+k} - \frac{k-1}{m+k-1}]} \big|W_h(\lambda)\big| \le \max_{0 \le \lambda \le 1/m} \big|W_h(\lambda)\big| \stackrel{\mathcal{D}}{=} \max_{0 \le \lambda \le 1} \big|W_h(\lambda)\big|/\sqrt{m} \,,$$

which yields

$$\mathbb{P}\Big(\big|\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) - \mathcal{W}_d(t_0)\big| > y_d\Big) = \mathbb{P}\Big(\max_{h=1}^{d} M_h(t_0) - \max_{h=1}^{d} M_{h,m}(t_0) > y_d\Big)$$

$$= \mathbb{P}\Big(\max_{h=1}^{d} \max_{t \in [t_0, T]} \max_{s \in [0, \, t-t_0]} \Big|W_h\Big(\frac{t}{t+1}\Big) - W_h\Big(\frac{s}{s+1}\Big)\Big|$$

$$- \max_{h=1}^{d} \max_{k=t_0 m+1}^{Tm} \max_{j=0}^{k-mt_0-1} \Big|W_h\Big(\frac{k}{m+k}\Big) - W_h\Big(\frac{j}{m+j}\Big)\Big| > y_d\Big)$$

$$\le \mathbb{P}\Big(\max_{h=1}^{d} \max_{k=t_0 m+1}^{Tm} \max_{j=0}^{k-mt_0-1} \max_{t \in [(k-1)/m, \, k/m]} \max_{s \in [j/m, \, (j+1)/m]} \Big|W_h(\tfrac{t}{t+1}) - W_h(\tfrac{s}{s+1})\Big|$$

$$- \max_{h=1}^{d} \max_{k=t_0 m+1}^{Tm} \max_{j=0}^{k-mt_0-1} \Big|W_h\Big(\frac{k/m}{1 + k/m}\Big) - W_h\Big(\frac{j/m}{1 + j/m}\Big)\Big| > y_d\Big)$$

$$\le \mathbb{P}\Big(\max_{h=1}^{d} \max_{k=t_0 m+1}^{Tm} \max_{j=0}^{k-mt_0-1} \max_{t \in [(k-1)/m, \, k/m]} \max_{s \in [j/m, \, (j+1)/m]}$$

$$\Big|W_h\Big(\frac{t}{t+1}\Big) - W_h\Big(\frac{s}{s+1}\Big) - W_h\Big(\frac{k/m}{1 + k/m}\Big) - W_h\Big(\frac{j/m}{1 + j/m}\Big)\Big| > y_d\Big)$$

$$\leq \mathbb{P}\left(2 \max_{h=1}^{d} \max_{k=1}^{Tm} \max_{t \in [(k-1)/m, \ k/m]} \left| W_h\left(\frac{t}{t+1}\right) - W_h\left(\frac{k/m}{1+k/m}\right) \right| > y_d \right)$$

$$\leq \sum_{h=1}^{d} \sum_{k=1}^{Tm} \mathbb{P}\left(\sup_{0 \leq \lambda \leq 1} |W_h(\lambda)| > y_d\sqrt{m}/2 \right) \leq dTm\frac{4}{\sqrt{2\pi}y_d}e^{-y_d^2 m/8} \ ,$$

where we have used the elementary bound [see for instance Karatzas and Shreve (1991)]

$$\mathbb{P}\left(\sup_{0 \leq \lambda \leq 1} |W_h(\lambda)| \geq z \right) \leq \frac{4}{\sqrt{2\pi}z}e^{-z^2/2} \ .$$

This yields (i) since by Assumption 3.1 (D1) the choice of y_d gives $my_d^2 = m^{1/3}$.

To obtain the estimate (ii), recall the definition of the Gaussian vector $\widetilde{\boldsymbol{V}}^{(z)} = \left(\widetilde{V}_1^{(z)}, \ldots, \widetilde{V}_{d_{\boldsymbol{V}}}^{(z)} \right)^{\top}$ in the proof of Lemma A.4, which yields the identity

$$\max_{i=1}^{d_{\boldsymbol{V}}} \widetilde{V}_i^{(z)} = \widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) \ .$$

For each component $\widetilde{V}_i^{(z)}$ of $\widetilde{\boldsymbol{V}}^{(z)}$ there are indices k, j, h such that

$$\widetilde{V}_i^{(z)} = \tilde{v}_{m,k,j,h} \ ,$$

where $\tilde{v}_{m,k,j,h}$ is s defined in (A.44). Thus, we obtain the following bounds for the variance of the components of $\widetilde{\boldsymbol{V}}^{(z)}$:

$$\text{Var}\left(\tilde{v}_{m,k,j,h} \right) = \frac{\left((k-j)w(k/m) \right)^2}{m} \, \text{Var}\left(\tilde{z}_{m+j+1}^{m+k}(h) - \tilde{z}_1^{m+j}(h) \right)$$

$$= \frac{\left((k-j)w(k/m) \right)^2}{m} \left[\text{Var}\left(\tilde{z}_{m+j+1}^{m+k}(h) \right) + \text{Var}\left(\tilde{z}_1^{m+j}(h) \right) \right]$$

$$\geq \frac{mt_0^2}{(1+T)^2} \, \text{Var}\left(\tilde{z}_1^{m+j}(h) \right) = \frac{mt_0^2}{(m+j)(1+T)^2} \geq \frac{t_0^2}{(1+T)^3}$$

and

$$\text{Var}\left(\tilde{v}_{m,k,j,h} \right) \leq \frac{mT^2}{(1+T)^2}\left(\frac{1}{k-j} + \frac{1}{m} \right) \leq \frac{T^2}{(1+T)^2}\frac{1+t_0}{t_0} \ .$$

Using these bounds, we can apply Lemma B.2 which yields

$$\mathbb{P}\left(\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) \leq u_d(x) \right) - \mathbb{P}\left(\widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) \leq u_d(x) - y_d \right)$$

$$= \mathbb{P}\left(-y_d \leq \widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) - u_d(x) \leq 0 \right) \leq \sup_{z \in \mathbb{R}} \mathbb{P}\left(\left| \widetilde{\mathcal{T}}_{m,d}^{(Z)}(t_0) - z \right| \leq y_d \right)$$

$$\leq C_{T,t_0} \cdot y_d\left(\sqrt{2\log(d)} + \sqrt{\max\{1, \log(\sigma_\ell/y_d)\}} \right) = o(1) \ ,$$

such that the assertion of Lemma A.5 follows by the choice of y_d. $\qquad\square$

Proof of Lemma A.6. First, recall the definition of $\mathcal{W}_d$ and $\mathcal{W}_d(t_0)$ in (A.18) and note that

$$\mathcal{W}_d = \max_{h=1}^{d} \max_{t \in [0,q(T)]} \max_{s \in [0,t]} \left| W_h(s) - W_h(t) \right|$$

$$= \max \left\{ \mathcal{W}_d(t_0) \,, \max_{h=1}^{d} \max_{t \in [0,q(t_0)]} \max_{s \in [0,t]} \left| W_h(s) - W_h(t) \right| \,, \right.$$

$$\left. \max_{h=1}^{d} \max_{t \in [q(t_0),q(T)]} \max_{s \in [q(q^{-1}(t)-t_0),t]} \left| W_h(s) - W_h(t) \right| \right\}$$

$$\leq \max \left\{ \mathcal{W}_d(t_0) \,, \max_{h=1}^{d} \max_{\substack{|t-s| \leq t_0 \\ s,t \in [0,q(T)]}} \left| W(t) - W(s) \right| \right\}$$

as $q(t_0) \leq t_0$ and $t - q(q^{-1}(t) - t_0) \leq t_0$. Hence, we obtain

$$\mathbb{P}\big(\mathcal{W}_d(t_0) \leq u_d(x)\big) - \mathbb{P}\big(\mathcal{W}_d \leq u_d(x)\big)$$

$$\leq \mathbb{P}\left(\max_{h=1}^{d} \max_{\substack{|t-s| \leq t_0 \\ s,t \in [0,q(T)]}} \left| W_h(t) - W_h(s) \right| > u_d(x) \right)$$

$$\leq d\, \mathbb{P}\left(\max_{\substack{|t-s| \leq t_0 \\ s,t \in [0,q(T)]}} \left| W_1(t) - W_1(s) \right| > u_d(x) \right) \,.$$

To control this probability we define an overlapping decomposition of the interval $[0, q(T)]$ by

$$I_j := [jt_0, (j+2)t_0] \,, \quad j = 0, 1, 2, \ldots, \lceil q(T)/t_0 \rceil - 2 \,.$$

Observing that the length of I_j is $2t_0$ we obtain

$$\mathbb{P}\left(\max_{\substack{|t-s| \leq t_0 \\ s,t \in [0,q(T)]}} \left| W_1(t) - W_1(s) \right| > u_d(x) \right)$$

$$\leq \sum_{j=1}^{\lceil q(T)/t_0 \rceil - 2} \mathbb{P}\left(\max_{\substack{|t-s| \leq t_0 \\ s,t \in I_j}} \left| W_1(t) - W_1(s) \right| > u_d(x) \right)$$

$$\leq \frac{q(T)}{t_0} \mathbb{P}\left(\max_{\substack{|t-s| \leq t_0 \\ s,t \in [0,2t_0]}} \left| W_1(t) - W_1(s) \right| > u_d(x) \right)$$

$$\leq \frac{q(T)}{t_0} \mathbb{P}\left(\max_{s,t \in [0,2t_0]} \left| W_1(t) - W_1(s) \right| > u_d(x) \right)$$

$$= \frac{q(T)}{t_0} \mathbb{P}\left(\max_{s,t \in [0,q(T)]} \left| W_1(t) - W_1(s) \right| > u_d(x) \cdot \sqrt{q(T)/(2t_0)} \right)$$

$$\leq \frac{q(T)}{t_0} \mathbb{P}\left(\max_{s,t \in [0,q(T)]} \left| W_1(t) - W_1(s) \right| > c\sqrt{\log(d)q(T)/(2t_0)} \right)$$

as for fixed x (3.6) yields, that there exists a positive constant $c < \sqrt{2q(T)}$, such that $u_d(x) \geq c \cdot \sqrt{\log(d)}$ for d sufficiently large. Using the representation of the distribution function $F_{\mathbb{M}}$ in (A.1) we obtain

$$\mathbb{P}\left(\max_{h=1}^{d} \max_{\substack{|t-s| \leq t_0 \\ s,t \in [0, q(T)]}} |W_h(t) - W_h(s)| > u_d(x) \right)$$

$$\leq d \frac{q(T)}{t_0} \left[1 - F_{\mathbb{M}}\left(c\sqrt{\log(d)q(T)/(2t_0)} \right) \right]$$

and L'Hôpital's rule gives

$$\lim_{d \to \infty} d \left[1 - F_{\mathbb{M}}\left(c\sqrt{\log(d)q(T)/(2t_0)} \right) \right]$$

$$= c\sqrt{q(T)/(2t_0)} \lim_{d \to \infty} d \frac{F_{\mathbb{M}}'\left(c\sqrt{\log(d)q(T)/(2t_0)} \right)}{2\sqrt{\log(d)}}$$

$$\leq c\sqrt{q(T)/(2t_0)} \lim_{d \to \infty} dF_{\mathbb{M}}'\left(c\sqrt{\log(d)q(T)/(2t_0)} \right).$$

Now substituting $d = \exp\left(\frac{2y^2 t_0}{c^2 q(T)} \right)$ yields that the last display can be written as

$$c\sqrt{q(T)/(2t_0)} \lim_{y \to \infty} \exp\left(\frac{2y^2 t_0}{c^2 q(T)} \right) F_{\mathbb{M}}'(y),$$

which by assertion (A.5) tends to zero for sufficiently small $t_0 > 0$ and thus completes the proof of Lemma A.6. $\qquad\square$

A.4. Proof of Theorem 3.10

Denote the size of the change by $\Delta\mu_m = |\mu_{m+k^*-1,h^*} - \mu_{m+k^*,h^*}|$ and the centered observations in component h^* by

$$X_{t,h^*}^{(c)} := X_{t,h^*} - \mathbb{E}[X_{t,h^*}].$$

Observe the following lower bound

$$\widehat{\mathcal{T}}_{m,d} = \max_{h=1}^{d} \max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h}(k) \overset{h=h^*}{\geq} \max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h^*}(k)$$

$$\overset{k=mT}{\geq} w(T)\hat{E}_{m,h^*}(mT) = \frac{1}{1+T} \max_{j=0}^{Tm-1} \frac{mT-j}{\sqrt{m}\hat{\sigma}_h} \left| \hat{\mu}_{m+j+1}^{m+mT}(h^*) - \hat{\mu}_1^{m+j}(h^*) \right|$$

$$\overset{j=k^*-1}{\geq} \frac{1}{1+T} \frac{mT-k^*+1}{\sqrt{m}\hat{\sigma}_h} \left| \hat{\mu}_{m+k^*}^{m+mT}(h^*) - \hat{\mu}_1^{m+k^*-1}(h^*) \right|$$

$$\geq \frac{1}{1+T} \left\{ \frac{mT-k^*+1}{\sqrt{m}\hat{\sigma}_h} \Delta\mu_m \right. \tag{A.49}$$

$$- \left| \frac{1}{\sqrt{m}\hat{\sigma}_h} \sum_{t=m+k^*}^{m+mT} X_{t,h^*}^{(c)} - \frac{mT - k^* + 1}{\sqrt{m}\hat{\sigma}_h(m + k^* - 1)} \sum_{t=1}^{m+k^*-1} X_{t,h^*}^{(c)} \right| \right\} .$$

The consistency of the long-run variance estimator $\hat{\sigma}_h$, Assumption 3.3, the FCLT in Theorem 3 of Wu (2005) and the Continuous Mapping Theorem show that

$$\left| \frac{1}{\sqrt{m}\hat{\sigma}_h} \sum_{t=m+k^*}^{m+mT} X_{t,h^*}^{(c)} - \frac{mT - k^* + 1}{\sqrt{m}\hat{\sigma}_h(m + k^* - 1)} \sum_{t=1}^{m+k^*-1} X_{t,h^*}^{(c)} \right|$$

$$\leq \max_{s \in [0,T]} \left| \frac{1}{\sqrt{m}\hat{\sigma}_h} \sum_{t=m+\lfloor ms \rfloor + 1}^{m+mT} X_{t,h^*}^{(c)} - \frac{mT - \lfloor ms \rfloor}{\sqrt{m}\hat{\sigma}_h(m + \lfloor ms \rfloor)} \sum_{t=1}^{m+\lfloor ms \rfloor} X_{t,h^*}^{(c)} \right|$$

$$\overset{\mathcal{D}}{\Longrightarrow} \max_{s \in [0,T]} \left| W(1 + T) - W(1 + s) - \frac{T - s}{1 + s} W(1 + s) \right| , \tag{A.50}$$

where W is a standard one-dimensional Brownian motion. Next, note that (A.49) gives that

$$\mathbb{P}\left(a_d \big(\widehat{\mathcal{T}}_{m,d} - b_d \big) > g_{1-\alpha} \right) = \mathbb{P}\left(\widehat{\mathcal{T}}_{m,d} > \frac{g_{1-\alpha}}{a_d} + b_d \right)$$

$$\geq \mathbb{P}\Bigg(- \left| \frac{1}{\sqrt{m}\hat{\sigma}_h} \sum_{t=m+k^*}^{m+mT} X_{t,h}^{(c)} - \frac{mT - k^* + 1}{\sqrt{m}\hat{\sigma}_h(m + k^* - 1)} \sum_{t=1}^{m+k^*-1} X_{t,h}^{(c)} \right| >$$

$$\left(\frac{g_{1-\alpha}}{a_d} + b_d \right)(T + 1) - \frac{mT - k^* + 1}{\sqrt{m}\hat{\sigma}_h} \Delta\mu_m \Bigg).$$

By Assumption 3.1, (D1) and (3.6) we obtain $b_d \lesssim \sqrt{\log m}$. Applying now (3.10) we get

$$\left(\frac{g_{1-\alpha}}{a_d} + b_d \right)(T + 1) - \frac{mT - k^* + 1}{\sqrt{m}\hat{\sigma}_h} \Delta\mu_m \overset{\mathbb{P}}{\Longrightarrow} -\infty . \tag{A.51}$$

Combining (A.50), (A.51) with an application of Slutsky's Theorem shows that the probability $\mathbb{P}\big(a_d \big(\widehat{\mathcal{T}}_{m,d} - b_d \big) > g_{1-\alpha} \big)$ tends to 1, which completes the proof.

A.5. Proof of Corollary 3.9

The result is obtained analogously to the corresponding parts of Theorem 2.5 in Jirak (2015b) or Theorem 3.11 in Dette and Gösmann (2018). Therefore the proof is omitted.

A.6. Proof of Theorem 3.14

Recall the definition of $u_d(x) = x/a_d + b_d$, $Z_{t,h}$, $\widehat{Z}_{t,h}$, $\mathcal{T}_{m,d}^{(Z)}$, $\mathcal{T}_{m,d}^{(Z)}(c)$ and $\widehat{\mathcal{T}}_{m,d}^{(Z)}(c)$ in (A.23), (A.19), (3.13), (A.20) and (A.22), respectively. The proof of Theorem 3.14 is based on the following three Lemmas.

Lemma A.7 *For the constant C_γ from Assumption 3.12 it holds that*

$$m^{C_\gamma} \max_{h,i=1}^{d} \left|\hat\rho_{h,i} - \rho_{h,i}\right| = o_\mathbb{P}(1) \ .$$

Proof. First, note that Lemma E.4 from Jirak (2015a) implies the existence of a global constant $C_\sigma > 0$, such that $\sup_{h\in\mathbb{N}} \sigma_h \leq C_\sigma$. Next, Assumption 3.3 and the Cauchy-Schwarz inequality imply that

(i) $\max_{h,i=1}^{d} \gamma_{h,i} \leq \sigma_h\sigma_i \leq C_\sigma^2$,

(ii) $\min_{h=1}^{d} \hat\sigma_h \geq \min_{h=1}^{d} \sigma_h - \max_{h=1}^{d}|\hat\sigma_h - \sigma_h| \geq c_\sigma - \max_{h=1}^{d}|\hat\sigma_h - \sigma_h|$,

(iii) $\max_{h,i=1}^{d} |\hat\sigma_h\hat\sigma_i - \sigma_h\sigma_i| \leq \max_{h,i=1}^{d} \hat\sigma_i|\hat\sigma_h - \sigma_h| + C_\sigma \max_{h=1}^{d}|\hat\sigma_h - \sigma_h|$
$$\leq C_\sigma \max_{h=1}^{d} |\hat\sigma_h - \sigma_h|^2 + 2C_\sigma \max_{h=1}^{d} |\hat\sigma_h - \sigma_h| \ .$$

Combining (i), (ii) and using again Assumption 3.3 gives

$$\max_{h,i=1}^{d} \left|\frac{\hat\gamma_{h,i}}{\hat\sigma_h\hat\sigma_i}\right| \leq \frac{1}{\left(c_\sigma - \max\limits_{h=1}^{d}|\hat\sigma_h - \sigma_h|\right)^2} \cdot \left(C_\sigma^2 + \max_{h,i=1}^{d}\left|\hat\gamma_{h,i} - \gamma_{h,i}\right|\right) = O_\mathbb{P}(1) \ .$$

Thus we obtain the upper bound

$$\max_{h,i=1}^{d} \left|\hat\rho_{h,i} - \rho_{h,i}\right| \leq \max_{h,i=1}^{d} \left|\frac{\hat\gamma_{h,i}}{\hat\sigma_h\hat\sigma_i} - \frac{\hat\gamma_{h,i}}{\sigma_h\sigma_i}\right| + \max_{h,i=1}^{d} \left|\frac{\hat\gamma_{h,i} - \gamma_{h,i}}{\sigma_h\sigma_i}\right|$$
$$\leq \frac{1}{c_\sigma^2} \max_{h,i=1}^{d} \left|\frac{\hat\gamma_{h,i}}{\hat\sigma_h\hat\sigma_i}\right| |\hat\sigma_h\hat\sigma_i - \sigma_h\sigma_i| + \frac{1}{c_\sigma^2} \max_{h,i=1}^{d} \left|\hat\gamma_{h,i} - \gamma_{h,i}\right|$$
$$\lesssim O_\mathbb{P}(1) \max_{h=1}^{d} |\hat\sigma_h - \sigma_h|^2 + O_\mathbb{P}(1) \max_{h=1}^{d} |\hat\sigma_h - \sigma_h| + \max_{h,i=1}^{d} \left|\hat\gamma_{h,i} - \gamma_{h,i}\right| \ .$$

The assertion of Lemma A.7 now follows from Assumption 3.12. $\qquad\square$

Lemma A.8 *There exists a sufficiently small constant $t_0 > 0$, such that for $x \in \mathbb{R}$ it holds*

$$\left|\mathbb{P}_{|\mathcal{X}}\left(\widehat{\mathcal{T}}_{m,d}^{(Z)}(t_0) \leq u_d(x)\right) - \mathbb{P}_{|\mathcal{X}}\left(\widehat{\mathcal{T}}_{m,d}^{(Z)} \leq u_d(x)\right)\right| = o_\mathbb{P}(1) , \qquad as \ m, d \to \infty \ .$$

Proof. We provide a (stochastic) version of the proof of Lemma A.2. First note that

$$\widehat{\mathcal{T}}_{m,d}^{(Z)} = \max_{h=1}^{d} \max_{k=1}^{Tm} \max_{j=0}^{k-1} \frac{(k-j)w(k/m)}{\sqrt{m}} \left|\widehat\mu_{m+j+1}^{m+k}(h) - \widehat\mu_1^{m+j}(h)\right|$$
$$= \max\left\{\widehat{\mathcal{T}}_{m,d}^{(Z)}(t_0), \max_{h=1}^{d} \max_{k=t_0m+1}^{Tm} \max_{j=k-t_0m}^{k-1} \frac{(k-j)w(k/m)}{\sqrt{m}} \left|\hat z_{m+j+1}^{m+k}(h) - \hat z_1^{m+j}(h)\right|, \right.$$
$$\left. \max_{h=1}^{d} \max_{k=1}^{t_0m} \max_{j=0}^{k-1} \frac{(k-j)w(k/m)}{\sqrt{m}} \left|\hat z_{m+j+1}^{m+k}(h) - \hat z_1^{m+j}(h)\right|\right\} .$$

Hence, we obtain

$$\left|\mathbb{P}_{|\mathcal{X}}\left(\widehat{\mathcal{T}}_{m,d}^{(Z)}(t_0) \leq u_d(x)\right) - \mathbb{P}_{|\mathcal{X}}\left(\widehat{\mathcal{T}}_{m,d}^{(Z)} \leq u_d(x)\right)\right| \leq P_1(x) + P_2(x) \ ,$$

where the random variables $P_1(x)$ and $P_2(x)$ are defined by

$$P_1(x) = \mathbb{P}_{|\mathcal{X}} \left(\max_{h=1}^{d} \max_{k=t_0 m+1}^{Tm} \max_{j=k-t_0 m}^{k-1} \frac{(k-j)}{\sqrt{m}} \left| \hat{z}_{m+j+1}^{m+k}(h) - \hat{z}_1^{m+j}(h) \right| \geq u_d(x) \right) ,$$

$$P_2(x) = \mathbb{P}_{|\mathcal{X}} \left(\max_{h=1}^{d} \max_{k=1}^{t_0 m} \max_{j=0}^{k-1} \frac{(k-j)}{\sqrt{m}} \left| \hat{z}_{m+j+1}^{m+k}(h) - \hat{z}_1^{m+j}(h) \right| \geq u_d(x) \right)$$

and we additionally used that $w(k/m) \leq 1$. To complete the proof, it suffices by Markov's inequality to establish that

$$\mathbb{E}[P_1(x)] = o(1) \quad \text{and} \quad \mathbb{E}[P_2(x)] = o(1) .$$

To prove these assertions, observe the bounds

$$\mathbb{E}[P_1(x)] = \mathbb{P} \left(\max_{h=1}^{d} \max_{k=t_0 m+1}^{Tm} \max_{j=k-t_0 m}^{k-1} \frac{(k-j)}{\sqrt{m}} \left| \hat{z}_{m+j+1}^{m+k}(h) - \hat{z}_1^{m+j}(h) \right| \geq u_d(x) \right)$$

$$\leq \sum_{h=1}^{d} \mathbb{P} \left(\max_{k=t_0 m+1}^{Tm} \max_{j=k-t_0 m}^{k-1} \frac{(k-j)}{\sqrt{m}} \left| \hat{z}_{m+j+1}^{m+k}(h) - \hat{z}_1^{m+j}(h) \right| \geq u_d(x) \right)$$

$$\text{(A.52)}$$

and

$$\mathbb{E}[P_2(x)] = \mathbb{P} \left(\max_{h=1}^{d} \max_{k=1}^{t_0 m} \max_{j=0}^{k-1} \frac{(k-j)}{\sqrt{m}} \left| \hat{z}_{m+j+1}^{m+k}(h) - \hat{z}_1^{m+j}(h) \right| \geq u_d(x) \right)$$

$$\leq \sum_{h=1}^{d} \mathbb{P} \left(\max_{k=1}^{t_0 m} \max_{j=0}^{k-1} \frac{(k-j)}{\sqrt{m}} \left| \hat{z}_{m+j+1}^{m+k}(h) - \hat{z}_1^{m+j}(h) \right| \geq u_d(x) \right) .$$

$$\text{(A.53)}$$

The terms in (A.52) and (A.53) can now be controlled by the same arguments as given in the proof of Lemma A.2. $\qquad \square$

Lemma A.9 *For $x \in \mathbb{R}$ it holds, as $m, d \to \infty$,*

$$\left| \mathbb{P}_{|\mathcal{X}} \left(\widehat{\mathcal{T}}_{m,d}^{(Z)} \leq u_d(x) \right) - \mathbb{P}_{H_0} \left(\mathcal{T}_{m,d} \leq u_d(x) \right) \right| = o_{\mathbb{P}}(1) . \qquad \text{(A.54)}$$

Proof. Observing Lemmas A.2, A.3, A.8, the assertion of Lemma A.9 follows, if we can establish that

$$\left| \mathbb{P}_{|\mathcal{X}} \left(\widehat{\mathcal{T}}_{m,d}^{(Z)}(t_0) \leq u_d(x) \right) - \mathbb{P} \left(\mathcal{T}_{m,d}^{(Z)}(t_0) \leq u_d(x) \right) \right| = o_{\mathbb{P}}(1) , \qquad \text{(A.55)}$$

To obtain this, we will reuse the vector technique applied in the proof of Lemma A.3. From the proof of this Lemma recall the definition and construction of the Gaussian vector $\boldsymbol{V}^{(z)} = \left(V_1^{(z)}, \ldots, V_{d_{\boldsymbol{V}}}^{(z)} \right)^{\top}$ which fulfilled the identity

$$\max_{i=1}^{d_{\boldsymbol{V}}} V_i^{(z)} = \mathcal{T}_{m,d}^{(Z)}(t_0) .$$

Analogously we construct a vector $\widehat{\boldsymbol{V}}^{(z)} = \left(\widehat{V}_1^{(z)}, \ldots, \widehat{V}_{d_{\boldsymbol{V}}}^{(z)} \right)^{\top}$ from $\left\{ \widehat{Z}_{t,h} \right\}$, such that

$$\max_{i=1}^{d_{\boldsymbol{V}}} \widehat{V}_i^{(z)} = \widehat{\mathcal{T}}_{m,d}^{(Z)}(t_0) \ .$$

The covariance structure of $\boldsymbol{V}^{(z)}$ was already calculated in Lemma A.3. Repeating these steps for the conditional covariance structure of $\widehat{\boldsymbol{V}}^{(z)}$ with respect to $\mathcal{X}$, we directly obtain that

$$\max_{i_1,i_2=1}^{d_{\boldsymbol{V}}} \left| \operatorname{Cov} \left(V_{i,1}^{(z)}, V_{i,2}^{(z)} \right) - \operatorname{Cov}_{|\mathcal{X}} \left(\widehat{V}_{i,1}^{(z)}, \widehat{V}_{i,2}^{(z)} \right) \right| \lesssim \max_{h,i=1}^{d} \left| \hat{\rho}_{h,i} - \rho_{h,i} \right|. \quad \text{(A.56)}$$

In the remainder of the proof we use the notation $\Delta_\rho = \max_{h,i=1}^{d} \left| \hat{\rho}_{h,i} - \rho_{h,i} \right|$. In view of (A.56), we are able to apply the Gaussian comparison inequality from Lemma B.3, which gives

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}_{|\mathcal{X}} \left(\widehat{\mathcal{T}}_{m,d}^{(Z)}(t_0) \leq x \right) - \mathbb{P} \left(\mathcal{T}_{m,d}^{(Z)} \leq x \right) \right|$$

$$= \sup_{x \in \mathbb{R}} \left| \mathbb{P}_{|\mathcal{X}} \left(\max_{i=1}^{d_{\boldsymbol{V}}} \widehat{V}_i^{(z)} \leq x \right) - \mathbb{P} \left(\max_{i=1}^{d_{\boldsymbol{V}}} V_i^{(z)} \leq x \right) \right|$$

$$\leq C \Delta_\rho^{1/3} \cdot \max \left\{ 1, \log \left(d_{\boldsymbol{V}} / \Delta_\rho \right) \right\}^{2/3}.$$

Due to Lemma A.7 and Assumption 3.1 the upper bound in the last display is of order $o_{\mathbb{P}}(1)$, which proves (A.55). □

Final step in proof of Theorem 3.14. To obtain the theorem's assertions, note that from Corollary 3.9 we already know that

$$a_d \left(\widehat{\mathcal{T}}_{m,d} - b_d \right) \stackrel{\mathcal{D}}{\Longrightarrow} G \ ,$$

and as the Gumbel distribution has a continuous c.d.f., Polya's theorem [see Serfling (2009), p. 18] directly implies convergence in Kolmogorov-metric, that is

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left(a_d \left(\widehat{\mathcal{T}}_{m,d} - b_d \right) \leq u_d(x) \right) - \mathbb{P} \left(G \leq x \right) \right| = o(1) \ . \quad \text{(A.57)}$$

On the other hand, combining (A.54) with Theorem 3.6 implies that

$$a_d \left(\widehat{\mathcal{T}}_{m,d}^{(Z)} - b_d \right) \stackrel{\mathcal{D}}{\Longrightarrow} G \ , \quad \text{(A.58)}$$

conditional on $\mathcal{X}$ in probability. So a conditional version of Polya's theorem gives

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}_{|\mathcal{X}} \left(a_d \left(\widehat{\mathcal{T}}_{m,d}^{(Z)} - b_d \right) \leq u_d(x) \right) - \mathbb{P} \left(G \leq x \right) \right| = o_{\mathbb{P}}(1) \ .$$

By (A.57) and (A.58) the proof of Theorem 3.14 is complete. □

A.7. Proof of Theorem 3.17

Denote the centered observations by

$$X_{t,h}^{(c)} = X_{t,h} - \mathbb{E}[X_{t,h}] \ .$$

We first prove assertions (3.16) and (3.18) using the Gumbel quantile $g_{1-\alpha}$.
   Proof of (3.16)**:** It holds that

$$
\begin{aligned}
\mathbb{P}\Big(\mathcal{S}_d \subset \widehat{\mathcal{S}}_{d,\alpha}\Big) &= \mathbb{P}\Big(\max_{h \in \mathcal{S}_d} \max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h}(k) \le g_{1-\alpha}/a_d + b_d \Big) \\
&= \mathbb{P}_{H_0}\Big(\max_{h \in \mathcal{S}_d} \max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h}(k) \le g_{1-\alpha}/a_d + b_d \Big) \\
&\ge \mathbb{P}_{H_0}\Big(\max_{h=1}^{d} \max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h}(k) \le g_{1-\alpha}/a_d + b_d \Big) \longrightarrow 1 - \alpha \ ,
\end{aligned}
$$

where we applied Corollary 3.9 for the last convergence.
   Proof of (3.18)**:** First, note that:

$$\mathbb{P}\Big(\mathcal{S}_d^c \subset \widehat{\mathcal{S}}_{d,\alpha}^c\Big) = \mathbb{P}\Big(\min_{h \in \mathcal{S}_d^c} \max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h}(k) > g_{1-\alpha}/a_d + b_d \Big) \ . \qquad \text{(A.59)}$$

We have the lower bound

$$
\begin{aligned}
\min_{h \in \mathcal{S}_d^c} \max_{k=1}^{Tm} w(k/m)\hat{E}_{m,h}(k) &\ge \min_{h \in \mathcal{S}_d^c} w(T)\hat{E}_{m,h}(Tm) \\
&\ge \min_{h \in \mathcal{S}_d^c} \frac{Tm - k_h^* - 1}{\sqrt{m}\hat{\sigma}_h(T+1)}\big|\widehat{\mu}_{m+k_h^*}^{m+Tm} - \widehat{\mu}_1^{m+k_h^*-1}\big| \ge A_1 - A_2 \ ,
\end{aligned}
$$

where the terms A_1 and A_2 are given by

$$A_1 = \min_{h \in \mathcal{S}_d^c} \frac{Tm - k_h^* + 1}{\sqrt{m}\hat{\sigma}_h(T+1)}\big|\mu_{m+k_h^*-1} - \mu_{m+k_h^*}\big| \ ,$$

$$A_2 = \max_{h \in \mathcal{S}_d^c} \frac{1}{\sqrt{m}\hat{\sigma}_h(T+1)}\Bigg| \sum_{t=m+k_h^*}^{m+Tm} X_{t,h}^{(c)} - \frac{mT - k_h^* + 1}{m + k_h^* - 1} \sum_{t=1}^{m+k_h^*-1} X_{t,h}^{(c)} \Bigg| \ .$$

Therefore the probability given in (A.59) has the lower bound

$$\mathbb{P}\big(a_d(A_1 - 2b_d) - a_d(A_2 - b_d) > g_{1-\alpha}\big) \ .$$

Using Corollary 3.9 we obtain

$$a_d\big(A_2 - b_d\big) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(A.60)}$$

$$\le a_d \Bigg(\max_{h=1}^{d} \max_{k=1}^{Tm} w(k/m) \max_{j=0}^{k-1} \Bigg| \sum_{t=m+j+1}^{m+k} X_{t,h}^{(c)} - \frac{k-j}{m+j} \sum_{t=1}^{m+j} X_{t,h}^{(c)} \Bigg| - b_d \Bigg) = O_{\mathbb{P}}(1) \ .$$

Further it holds by (3.17) that for m sufficiently large

$$Tm - \max_{h \in \mathcal{S}_d^c} k_h^* > c \ ,$$

where $c > 0$ is a sufficiently small constant. By Assumptions 3.3 and 3.12 we have

$$\frac{1}{\max\limits_{h=1}^{d} \hat{\sigma}_h} \geq \frac{1}{C_\sigma + \max\limits_{h=1}^{d} |\hat{\sigma}_h - \sigma_h|} \overset{\mathbb{P}}{\Longrightarrow} \frac{1}{C_\sigma}$$

and (3.6) shows that $b_d \lesssim \sqrt{\log m}$ and $a_d \to \infty$. Combining this with the assertions above yields

$$a_d\big(A_1 - 2b_d\big) \tag{A.61}$$
$$\gtrsim a_d\left(\sqrt{m}\frac{1}{\max\limits_{h=1}^{d} \hat{\sigma}_h} \min_{h \in \mathcal{S}_d^c} \big|\mu_{m+k_h^*-1} - \mu_{m+k_h^*}\big| - 2b_d\right) \overset{\mathbb{P}}{\Longrightarrow} \infty.$$

A combination of (A.60) and (A.61) now proves (3.18).

To complete the proof of Theorem 3.17 it remains to discuss the case, where bootstrap quantiles

$$\hat{q}_{m,1-\alpha} := \inf\left\{x \in \mathbb{R} \;\middle|\; \mathbb{P}_{|\mathcal{X}}\left(a_d\big(\widehat{\mathcal{T}}_{m,d}^{(Z)} - b_d\big) \leq x\right) \geq \alpha\right\},$$

are used in the algorithm. Fortunately, it follows from Theorem 3.14 combined with Lemma 21.2 and (the arguments from) Lemma 23.2 in van der Vaart (1998) that

$$\hat{q}_{m,1-\alpha} \overset{\mathbb{P}}{\Longrightarrow} g_{1-\alpha}.$$

An application of Slutsky's Lemma to the statements above then completes the proof of Theorem 3.17.

Appendix B: Technical auxiliary results

We require the following Nagaev-type inequality as given in the online supplement of Jirak (2015b) which is a version of Theorem 2 in Liu et al. (2013). In particular the reader should note that the bound is independent of h.

Lemma B.1 *Under Assumption 3.3 it holds for* $x \geq C_p'\sqrt{n}$

$$\mathbb{P}\left(\max_{k=1}^{n}\left|\sum_{t=1}^{k} X_{t,h} - \mathbb{E}[X_{t,h}]\right| > x\right) \leq C_p\frac{n}{x^p} + C_p\exp\left(-c_p\frac{x^2}{n}\right),$$

where the constants $c_p, C_p, C_p > 0$ *depend on* p *and the sequence* $\left\{\sup_{h\in\mathbb{N}} \vartheta_{t,h,p}\right\}_{t\in\mathbb{N}}$ *only.*

As an immediate consequence of the bound

$$\max_{k=1}^{n}\left|\sum_{t=k}^{n} X_{t,h} - \mathbb{E}[X_{t,h}]\right| \leq 2\max_{k=1}^{n}\left|\sum_{t=1}^{k} X_{t,h} - \mathbb{E}[X_{t,h}]\right|,$$

Lemma B.1 holds with adjusted constants also for the reversed partial sum maximum.

The following inequality is Lemma 2.1 in Chernozhukov et al. (2013).

Lemma B.2 *Let $\boldsymbol{Z} = (Z_1, \ldots, Z_d)^\top$ be a centered Gaussian vector with covariance matrix Σ^Z whose diagonal entries are bounded by two constants σ_ℓ and σ_u, that is*

$$0 < \sigma_\ell \leq \Sigma^Z_{j,j} \leq \sigma_u$$

for $j = 1, \ldots, d$. Then for $\delta > 0$ it holds that

$$\sup_{z \in \mathbb{R}} \mathbb{P}\Big(\Big| \max_{h=1}^{d} Z_h - z \Big| \leq \delta \Big) \leq C_\sigma \delta \Big(\sqrt{2\log(d)} + \sqrt{\max\{1, \log(\sigma_\ell/\delta)\}} \Big) \ ,$$

where the constant $C_\sigma > 0$ depends on σ_ℓ and σ_u only.

The next tool is Lemma 3.1 from Chernozhukov et al. (2013).

Lemma B.3 *Let $\boldsymbol{U} = (U_1, \ldots, U_d)^\top$ and $\boldsymbol{V} = (V_1, \ldots, V_d)^\top$ denote two centered, d-dimensional Gaussian vectors with covariance matrices Σ^U and Σ^V, respectively. Further assume that there are two constants $c_1, C_1 > 0$, such that for all $j = 1, \ldots, d$*

$$c_1 \leq \big| \Sigma^U_{j,j} \big| \leq C_1 \ .$$

Denote the maximum entry-wise distance of both covariance matrices by

$$\Delta := \max_{i,j=1}^{d} \big| \Sigma^U_{i,j} - \Sigma^V_{i,j} \big| \ .$$

Then it holds that

$$\sup_{x \in \mathbb{R}} \Big| \mathbb{P}\Big(\max_{i=1}^{d} U_i \leq x \Big) - \mathbb{P}\Big(\max_{i=1}^{d} V_i \leq x \Big) \Big| \leq C\Delta^{1/3} \cdot \max\Big\{ 1, \log\big(d/\Delta\big) \Big\}^{2/3} \ ,$$

where the constant $C > 0$ depends on c_1 and C_1 only.

Acknowledgments

References

Anatolyev, S. and Kosenok, G. (2018). Sequential testing with uniformly distributed size. *Journal of Time Series Econometrics*, 10(2). MR3833440

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858. MR1106513

Arratia, R., Goldstein, L., and Gordon, L. (1989). Two moments suffice for Poisson approximations: The Chen-Stein method. *The Annals of Probability*, 17(1):9–25. MR0972770

Aue, A., Hörmann, S., Horváth, L., Hušková, M., and Steinebach, J. G. (2012). Sequential testing for the stability of high-frequency portfolio betas. *Econometric Theory*, 28(4):804–837. MR2959126

Aue, A. and Horváth, L. (2004). Delay time in sequential detection of change. *Statistics & Probability Letters*, 67(3):221–231. MR2053524

Aue, A., Horváth, L., Hušková, M., and Kokoszka, P. (2006). Change-point monitoring in linear models. *The Econometrics Journal*, 9(3):373–403. MR2328569

Avanesov, V. and Buzun, N. (2018). Change-point detection in high-dimensional covariance structure. *Electronic Journal of Statistics*, 12(2):3254–3294. MR3861282

Berkes, I., Liu, W., and Wu, W. B. (2014). Komlós-Major-Tusnády approximation under dependence. *The Annals of Probability*, 42(2):794–817. MR3178474

Billingsley, P. (1999). *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, second edition. MR1700749

Borodin, A. and Salminen, P. (1996). *Handbook of Brownian Motion - Facts and Formulae*. Probability and its applications. Birkhäuser, Basel. MR1477407

Chen, Y., Wang, T., and Samworth, R. J. (2020). High-dimensional, multiscale online changepoint detection. *arXiv e-print arXiv:2003.03668v1*. MR4400396

Chen, Z. and Tian, Z. (2010). Modified procedures for change point monitoring in linear models. *Mathematics and Computers in Simulation*, 81(1):62–75. MR2729128

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6):2786–2819. MR3161448

Chernozhukov, V., Chetverikov, D., Kato, K., and Koike, Y. (2019). Improved central limit theorem and bootstrap approximations in high dimensions. *arXiv e-print arXiv:1912.10529v1*. MR3693963

Cho, H. and Fryzlewicz, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(2):475–507. MR3310536

Chochola, O., Hušková, M., Prášková, Z., and Steinebach, J. G. (2013). Robust monitoring of capm portfolio betas. *Journal of Multivariate Analysis*, 115:374–395. MR3004565

Chu, C.-S. J., Stinchcombe, M., and White, H. (1996). Monitoring structural change. *Econometrica*, 64(5):1045–1065.

Chu, L. and Chen, H. (2018). Sequential change-point detection for high-dimensional and non-euclidean data. *arXiv e-print arXiv:1810.05973v1*. MR4035309

Ciuperca, G. (2013). Two tests for sequential detection of a change-point in a

nonlinear model. *Journal of Statistical Planning and Inference*, 143(10):1719–1743. MR3082229

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer Science+Business Media, New York. MR2234156

Dette, H., Bornkamp, B., and Bretz, F. (2013). On the efficiency of two-stage response-adaptive designs. *Statistics in Medicine*, 32(10):1646–1660. MR3060632

Dette, H. and Gösmann, J. (2018). Relevant change points in high dimensional time series. *Electronic Journal of Statistics*, 12(2):2578–2636. MR3849896

Dette, H. and Gösmann, J. (2019). A likelihood ratio approach to sequential change point detection for a general class of parameters. *Journal of the American Statistical Association*. Preprint available at: https://doi.org/10.1080/01621459.2019.1630562. MR4143471

Dette, H., Pan, G. M., and Yang, Q. (2018). Estimating a change point in a sequence of very high-dimensional covariance matrices. *arXiv e-print arXiv:1807.10797v1*. MR4399097

El Machkouri, M., Volný, D., and Wu, W. B. (2013). A central limit theorem for stationary random fields. *Stochastic Processes and their Applications*, 123(1):1–14. MR2988107

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, volume 33 of *Applications of Mathematics (New York)*. Springer, Berlin. MR1458613

Enikeeva, F. and Harchaoui, Z. (2019). High-dimensional change-point detection under sparse alternatives. *The Annals of Statistics*, 47(4):2051–2079. MR3953444

Fan, J., Liao, Y., and Yao, J. (2015). Power enhancement in high-dimensional cross-sectional tests. *Econometrica*, 83(4):1497–1541. MR3384226

Feller, W. (1951). The asymptotic distribution of the range of sums of independent random variables. *The Annals of Mathematical Statistics*, 22(3):427–432. MR0042626

Fremdt, S. (2014). Asymptotic distribution of the delay time in Page's sequential procedure. *Journal of Statistical Planning and Inference*, 145:74–91. MR3125350

Fremdt, S. (2015). Page's sequential procedure for change-point detection in time series regression. *Statistics*, 49(1):128–155. MR3304372

Gösmann, J., Kley, T., and Dette, H. (2020). A new approach for open-end sequential change point monitoring. *To appear in: Journal of Time Series Analysis*. Preprint available at https://doi.org/10.1111/jtsa.12555.

Hawkins, D. M., Qiu, P., and Kang, C. W. (2003). The changepoint model for statistical process control. *Journal of Quality Technology*, 35(4):355–366.

Hinkley, D. V. (1971). Inference about the change-point from cumulative sum tests. *Biometrika*, 58(3):509–523. MR0312623

Hoga, Y. (2017). Monitoring multivariate time series. *Journal of Multivariate Analysis*, 155:105–121. MR3607884

Horváth, L., Hušková, M., Kokoszka, P., and Steinebach, J. (2004). Monitor-

ing changes in linear models. *Journal of Statistical Planning and Inference*, 126(1):225–251. MR2090695

Hušková, M. and Kirch, C. (2012). Bootstrapping sequential change-point tests for linear regression. *Metrika*, 75(5):673–708. MR2946654

Hušková, M. and Koubková, A. (2005). Monitoring jump changes in linear models. *Journal of Statistical Research*, 39(2):51–70. MR2256552

Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *The Annals of Applied Probability*, 14(2):865–880. MR2052906

Jirak, M. (2015a). Supplement to: Uniform change point tests in high dimension. *The Annals of Statistics*, 43(6):2451–2483. Supplement availabe at: https://doi.org/10.1214/15-AOS1347SUPP.

Jirak, M. (2015b). Uniform change point tests in high dimension. *The Annals of Statistics*, 43(6):2451–2483. MR3405600

Karatzas, I. and Shreve, S. (1991). *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition. MR1121940

Kaul, A., Jandhyala, V. K., and Fotopoulos, S. B. (2019). An efficient two step algorithm for high dimensional change point regression models without grid search. *J. Mach. Learn. Res.*, 20:Paper No. 111, 40. MR3990465

Kirch, C. (2008). Bootstrapping sequential change-point tests. *Sequential Analysis*, 27(3):330–349. MR2446906

Kirch, C. and Kamgaing, J. T. (2015). On the use of estimating functions in monitoring time series for change points. *Journal of Statistical Planning and Inference*, 161:25–49. MR3316549

Kirch, C. and Stoehr, C. (2019). Sequential change point tests based on U-statistics. *arXiv e-print arXiv:1912.08580v1*. MR2446906

Kirch, C. and Weber, S. (2018). Modified sequential change point procedures based on estimating functions. *Electronic Journal of Statistics*, 12(1):1579–1613. MR3806433

Kock, A. B. and Preinerstorfer, D. (2019). Power in high-dimensional testing problems. *Econometrica*, 87(3):1055–1069. MR3957338

Lai, T. L. (2001). Sequential analysis: Some classical problems and new challenges. *Statistica Sinica*, 11(2):303–351. MR1844531

Lévy-Leduc, C. and Roueff, F. (2009). Detection and localization of change-points in high-dimensional network traffic data. *Annals of Applied Statistics*, 3(2):637–662. MR2750676

Liu, W., Xiao, H., and Wu, W. B. (2013). Probability and moment inequalities under dependence. *Statistica Sinica*, 23(3):1257–1272. MR3114713

Mei, Y. (2008). Is average run length to false alarm always an informative criterion? *Sequential Analysis*, 27(4):354–376. MR2460205

Mei, Y. (2010). Efficient scalable schemes for monitoring a large number of data streams. *Biometrika*, 97(2):419–433. MR2650748

Moustakides, G. V. (1986). Optimal stopping times for detecting changes in distributions. *The Annals of Statistics*, 14(4):1379–1387. MR0868306

Nikiforov, I. (1987). Sequential detection of changes in stochastic systems. *IFAC*

Proceedings Volumes, 20(2):321–327.

Otto, S. and Breitung, J. (2019). Backward CUSUM for testing and monitoring structural change. *arXiv e-print arXiv:2003.02682v1*.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2):100–115. MR0088850

Page, E. S. (1955). Control charts with warning lines. *Biometrika*, 42(1/2):243–257. MR0072412

Ross, G. J. (2014). Sequential change detection in the presence of unknown parameters. *Statistics and Computing*, 24(6):1017–1030. MR3253851

Schröter, K., Mühr, B., Elmer, F., Kunz-Plapp, T., and Trieselmann, W. (2013). June 2013 flood in central Europe - focus Germany. *CEDIM Forensic Disaster Analysis Group (FDA)*. https://www.cedim.kit.edu/2850.php.

Serfling, R. J. (2009). *Approximation Theorems of Mathematical Statistics. Wiley Series in Probability and Statistics.* John Wiley & Sons, Inc., New York. MR0595165

Sharipov, O., Tewes, J., and Wendler, M. (2016). Sequential block bootstrap in a Hilbert space with application to change point analysis. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 44(3):300–322. MR3536199

Soh, Y. S. and Chandrasekaran, V. (2017). High-dimensional change-point estimation: Combining filtering with convex optimization. *Applied and Computational Harmonic Analysis*, 43(1):122–147. MR3638952

Steland, A. (2006). A bootstrap view on Dickey-Fuller control charts for AR(1) series. *Austrian Journal of Statistics*, 35:339–346.

Tartakovsky, A., Nikiforov, I. V., and Basseville, M. (2014). *Sequential Analysis: Hypothesis Testing and Changepoint Detection*, volume 136 of *Chapman & Hall/CRC Monographs on Statistics & Applied Probability*. Chapman & Hall/CRC, Taylor and Francis Group. MR3241619

Tartakovsky, A. G., Rozovskii, B. L., Blazek, R. B., and Hongjoong Kim (2006). A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods. *IEEE Transactions on Signal Processing*, 54(9):3372–3382. MR2240956

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge. MR1652247

Wald, A. (1945). Sequential tests of statistical hypotheses. *Annals of Mathematical Statistics*, 16(2):117–186. MR0013275

Wang, D., Yu, Y., and Rinaldo, A. (2017). Optimal covariance change point localization in high dimension. *arXiv e-print arXiv:1712.09912v1*. MR4177380

Wang, D., Zhao, Z., Lin, K. Z., and Willett, R. (2021). Statistically and computationally efficient change point localization in regression settings. *J. Mach. Learn. Res.*, 22:Paper No. [248], 46. MR4353027

Wang, R. and Shao, X. (2020). Dating the break in high-dimensional data. *arXiv e-print arXiv:2002.04115v1*.

Wang, R., Volgushev, S., and Shao, X. (2019). Inference for change points in high dimensional data. *arXiv e-print arXiv:1905.08446v1*.

Wang, T. and Samworth, R. J. (2018). High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):57–83. MR3744712

Wied, D. and Galeano, P. (2013). Monitoring correlation change in a sequence of random variables. *Journal of Statistical Planning and Inference*, 143(1):186–196. MR2969021

Woodall, W. H. and Montgomery, D. C. (1999). Research issues and ideas in statistical process control. *Journal of Quality Technology*, 31(4):376–386.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America*, 102(40):14150–14154. MR2172215

Wu, W. B. and Zhou, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statistica Sinica*, 21(3):1397–1413. MR2827528

Xie, Y. and Siegmund, D. (2013). Sequential multi-sensor change-point detection. *The Annals of Statistics*, 41(2):670–692. MR3099117

Yu, Y., Padilla, O. H. M., Wang, D., and Rinaldo, A. (2020). A note on online change point detection. *arXiv e-print arXiv:2006.03283v1*. MR4395506

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, 11(10):1–17.

Zhang, X. and Cheng, G. (2018). Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli*, 24(4A):2640–2675. MR3779697

Zou, C., Wang, Z., Zi, X., and Jiang, W. (2015). An efficient online monitoring method for high-dimensional data streams. *Technometrics*, 57(3):374–387. MR3384952